Summer 7-2012

MECHANISMS OF GROUP II INTRON LOSS FROM THE MITOCHONDRIAL *COX2* GENE OF PLANTS

Nancy J. Hepburn
University of Nebraska-Lincoln

MECHANISMS OF GROUP II INTRON LOSS FROM THE MITOCHONDRIAL *COX2* GENE OF PLANTS

By

Nancy Jimenez Hepburn

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Agronomy

Under the Supervision of Professor Jeffrey P. Mower

Lincoln, Nebraska

July, 2012

MECHANISMS OF GROUP II INTRON LOSS FROM THE MITOCHONDRIAL *COX2* GENE OF PLANTS

Nancy Jimenez Hepburn, M.S.
University of Nebraska, 2012

Adviser: Jeffrey P. Mower

It has been 35 years since the discovery of introns. Although fundamental aspects of their function and evolutionary origins have been widely studied, the mechanisms of intron loss are still an unresolved evolutionary mystery. Previously proposed mechanisms of intron loss include exonization, random genomic deletion and retroprocessing. The mitochondrial genomes of land plants have been characterized by widespread intron loss. In angiosperms, two introns located in the *cox2* gene display a particularly variable intron distribution, indicating frequent loss. We took advantage of the frequent intron loss from *cox2* to investigate the possible mechanisms of loss from this gene. The first study focused on *Maganolia tripetala*, a species that contains two *cox2* gene copies that differ in intron content. Transcriptional and phylogenetic analyses indicated that the edit site distribution was not consistent with the retroprocessing model. Instead, we showed that the loss of two group II introns from the mitochondrial *cox2* gene of *Magnolia tripetala* was mediated by a novel mechanism involving horizontal gene transfer and gene conversion (HGT-GC). In a broader study involving all angiosperms, we found no support for intron loss via exonization and random genomic deletion. We also did not find strong evidence supporting retroprocessing as the sole mechanism of intron loss. Our overall findings suggest that HGT-GC or an unidentified mechanism plays a larger role in the loss of introns than previously recognized.

TABLE OF CONTENTS

CHAPTER 3:

Mechanisms of group II intron loss from the mitochondrial *cox2* gene of plants

ACKNOWLEDGEMENTS

First, I would like to thank God, because through him, everything is possible (including this degree). I would also like to express my appreciation and respect to Dr. Jeffrey P. Mower for his great patience and his understanding, for sharing his knowledge with me, for his guidance, and for being an outstanding boss and amazing mentor. I would like to recognize Derek W. Schmidt for his friendship and his help with both *cox2*-intron related projects.

I am very thankful to Dr. Thomas Elthon, Dr. Milford Hanna, and Dr. George Graef for being amazing professors, teachers and ex-bosses. Also, I am obliged to Dr. Alan Christensen and Dr. Stacey Smith for serving on my oral examining committee.

I am especially grateful to my husband Christopher, my parents Carmen and Enrique and my dog Mongo, because they encouraged me when life was difficult and they provided me with unconditional love.

Finally, I am thankful to my lab mates for their camaraderie and for making the Mower Lab a fun place to be, to the Mackenzie Lab for their helpful discussions, and to the Department of Agronomy and Horticulture for this wonderful opportunity.

CHAPTER 1

Introduction to plant mitochondria and intron loss

ABSTRACT

The mitochondrial genomes of land plants are complex when compared to their animal counterparts. While animal mitochondrial genomes are circular and relatively small (16-18 kb), the mitochondrial genomes of tracheophytes are dynamic and exist in linear and circular configurations with genome sizes ranging from 200 kb-11.3 Mb. Some of the factors known to contribute to the genome expansion and added complexity are: extensive recombination between molecules, the presence of group I and II introns, intracellular gene transfer and horizontal gene transfer. Another peculiarity of plant mitochondrial genes is RNA editing, which is a process that corrects mitochondrial transcripts. Here, we review some of these concepts, which elucidate why plant mitochondria are the ideal system to study mechanisms of intron loss (Chapters 2 and 3).

PLANT MITOCHONDRIAL GENOMES

Mitochondria are considered the power houses of the eukaryotic cells. In animal cells they are considered the main source of ATP. The main role of mitochondria is to oxidize carbohydrates, lipids and amino acids to ultimately obtain energy in the form of ATP. Thus, increasing our understanding of all the processes related to mitochondria is important because of its crucial role in animal and plant cells (Burger, Gray and Franz Lang 2003; Logan 2006).

It has been well established by the scientific community that the mitochondrion originated from a free living, aerobic and heterotrophic α-proteobacterium that became engulfed and subsequently taken over by a proto-ekaryote cell. Over time, most of its genes were transferred to the nucleus through a process known as intracellular gene transfer (IGT). As a consequence of IGT, mitochondria depend heavily on the nucleus for supplying most of its proteins and regulating all of its remaining genes. Mitochondral gene transfer to the nucleus is still an ongoing process in plants but not in most animals or fungi, in which the genomes have stabilized (Lang, Gray and Burger 1999; Richardson and Palmer 2007).

The mitochondrial genome varies greatly between animals and plants (Knoop et al. 2011). While animal mitochondrial genomes are circular and relatively small (roughly about 16 kb), plant mitochondrial genomes are highly variable in size (even between closely related species) and exist in two configurations: linear and circular (Alverson et al. 2010; Mower, Sloan and Alverson 2012). The genomes in a plant cell can be found in multiple configurations that differ in genome content, gene order, and sizes of repeats. It is through the presence of these repeats (especially large repeats) that plant mitochondria are capable of undergoing homologous recombination between the

different configurations (Lonsdale et al. 1988; Buchanan, Gruissem and Jones 2000). Moreover, the various molecules can be found at different stoichiometry that can change over time through inter- and intra-molecular recombination and self-replication also known as substoichiometric shifting (Lonsdale et al. 1988; Janska and Mackenzie 1993; Janska et al. 1998). One theory suggests that this is a mitochondrial strategy to deal with arising mutations (Maréchal and Brisson 2010). In other words, a mutation arising in a particular genome could be reversed by undergoing homologous recombination with a normal molecule. Another theory proposes that this division helps with the maintenance of phenotypic variation, or even maintains some of the dominant negative mutations at low levels (Small, Suffolk and Leaver 1989; Mackenzie and McIntosh 1999; Buchanan, Gruissem and Jones 2000). It is likely that this kind of recombination activity has caused great variability in genome synteny across species.

Plant mitochondrial genomes are more variable than the genomes of most other organisms. For instance, in tracheophytes, the mitochondrial genomes range from 200 kb to 11.3 Mb (Mower, Sloan and Alverson 2012; Sloan et al. 2012). It seems that in higher plants, most of the differences in genome sizes are due to intergenic (non-coding, featureless) DNA (Mower, Sloan and Alverson 2012). Multiple lines of evidence indicate that sequences from the nucleus and chloroplasts have been integrated into the mitochondrial genome via IGT (Sloan et al. 2010; Alverson et al. 2011; Mower, Sloan and Alverson 2012). Another factor influencing the genome sizes in plant mitochondria is the integration of foreign DNA from other species (inter specific gene transfer) as well as the insertion of group I and group II introns (Mower et al, in press). This takes us to horizontal gene transfer (HGT), also known as lateral transfer, which is a very common phenomenon in plant mitochondria (Bergthorsson et al. 2003; Woloszynska et al. 2004; Keeling and Palmer 2008). The exact mechanism by which foreign DNA is horizontally

transferred between plants is still unknown, but illegitimate pollination has been suggested to play an important role as well as vectoring agents such as mycorhizal fungi, pathogens and insects (Keeling and Palmer 2008).

PLANT MITOCHONDRIAL GROUP I AND II INTRONS

As previously mentioned, plant mitochondria genes are often interrupted by different types of mobile genetic elements, including group I and II introns. Group I introns are highly mobile homing ribozymes that possess a site-specific endonuclease, which is used to invade intronless alleles by means of the double-strand break repair pathway. In other words, a double-strand break is made in the DNA of the recipient allele followed by exonucleolytic degradation of the cleaved target sequence(s). Then, recombination occurs between the recipient and the donor (the exonic sequences flanking the intron). In this manner the intron-containing allele successfully transfers its intron and portions of the flanking exons. This process is followed by gap repair using the donor sequences as templates forming co-conversion tracts (CCT's) (Vaughn et al. 1995; Mueller, Smith and Belfort 1996; Lambowitz et al. 1999; Sanchez-Puerta et al. 2011).

Generally speaking, group I introns are not widely spread across land plants. One exception is the intron located within the mitochondrial *cox1* gene in many angiosperms (Cho et al. 1998). It is believed that this intronic sequence was initially horizontally acquired from a fungal source and subsequently profusely distributed across angiosperms via HGT. Furthermore, there are multiple lines of evidence from different researchers pointing towards HGT as the mechanism at work (Cho et al. 1998; Barkman et al. 2007; Sanchez-Puerta et al. 2008; Sanchez-Puerta et al. 2011). First, the *cox1* intron has a patchy distribution across angiosperms. Second, species that possess the intron also have highly divergent CCT's. Third, for the most part, intron phylogeny does

not resemble established organismal relationships. The most recent phylogenetic analyses from Sanchez-Puerta et al. (2011) corroborates previous studies by showing once again that many of the angiosperm *cox1* intron relationships are incongruent.

Group II introns, on the other hand, are the most abundant type of intron in plant mitochondria (Bonen and Vogel 2001). They are self-splicing ribozymes mainly composed of two parts: a ribozyme and an intron-encoded protein (IEP) forming a ribonucleic protein (Lambowitz and Zimmerly 2004; Lambowitz and Zimmerly 2011). The IEP can be further subdivided into four active domains: reverse transcriptase domain (RT), maturase domain (X), DNA binding domain and an endonuclease domain. Structurally speaking, group II introns have conserved secondary structures that encompass 6 helical domains labeled as (I-VI) that extend out from a main wheel (Lambowitz and Zimmerly 2004).

Group II introns are found mainly in eubacterial genomes (and to a lesser extent in archaea) as well as in the mitochondrial genomes of plants, fungi and protists. They are also present in the chloroplast but not in the nuclear genome of any eukaryote (Bonen 2008; Bonen 2011). Group II introns are further subdivided into eight more sub-lineages based on the IEP associations with the specific ribozymes (Simon, Kelchner and Zimmerly 2009). Curiously, bacterial group II introns encompass all sub-lineages while organellar lineages are restricted to the mitochondria-like lineage (ML) and chloroplast-like (CL) lineage. This phylogenetic distribution suggests that they were originally transferred to the eukaryote cell through the endosymbiosis events that gave rise to chloroplasts and mitochondria (Lambowitz and Zimmerly 2004; Simon, Kelchner and Zimmerly 2009; Lambowitz and Zimmerly 2011).

In early plant evolution, group II introns were able to rampantly spread within and among diverse mitochondrial genomes. At some point, however, plants managed to stop their spread by disrupting theRT domain located within the IEP. In other words, mutations are often found within the RT domains that knock out the RT capability of otherwise functional IEP's. It turns out that an intact maturase domain is not only essential for spreading but also for splicing (Lambowitz and Zimmerly 2004; Lambowitz and Zimmerly 2011). Fortunately, plants have figured out ways to deal with the introns that interrupt their coding regions. It seems that some introns with intact IEP ORF (and defective RT-domains) are able to recognize certain structural features of different introns (located within the same organelle) and assist with their splicing (Lambowitz and Zimmerly 2004; Lambowitz and Zimmerly 2011). This is the case for the *matR* gene in plant mitochondria, and the *matK* gene in the chloroplast. These maturases, with the assistance of nuclear encoded proteins (such as PPR and nMat proteins, among others), help fold and/or stabilize the intron into its catalytically active form for efficient splicing (Lambowitz and Zimmerly 2011).

RNA EDITING IN PLANT MITOCHONDRIA

Another peculiarity of plant mitochondrial and chloroplast genomes is RNA editing. RNA editing is a post-transcriptional process that occurs during RNA maturation. It leads to the conversion of specific genome-encoded cytosines to uracils (C to U), and with a lesser extent, the conversion of genome-encoded thymines to cytosines (U to C) (Shikanai 2006; Bruhs and Kempken 2011). Although most of the editing tends to occur within coding regions, it has also been observed to occur in introns, rRNA, and tRNA (Gray and Covello 1993; Mower 2008). Since most of the RNA edit sites appear to occur in codons within the first and second nucleotide positions (which changes the encoded amino acid), RNA editing seems to play a crucial role in correcting for harmful

mitochondrial genome mutations that ultimately would affect protein function (Shikanai 2006; Bruhs and Kempken 2011). In introns, it can also play a very important role because RNA editing corrects mis-pairings at critical positions in the predicted secondary structure of the intron‘s lariat (interaction between bases). This, in turn, helps with intron folding, which is ultimately necessary for intron splicing and transcript maturation (Börner et al. 1995; Li-Pook-Than et al. 2007).

MECHANISMS OF INTRON LOSS

Intron distribution across eukaryotes is highly variable, ranging from two in the genome of *Giardia lamblia* to more than 100,000 in many vertebrates and plants (Roy and Gilbert 2006; Li et al. 2009). It is likely that differences in intron content reflect different needs and selective pressures among organisms (Majewski and Ott 2002; Coulombe-Huntington and Majewski 2007b).

Most studies indicate that eukaryotic evolution has been primarily characterized by intron losses rather than gains (Mourier and Jeffares 2003; Roy, Fedorov and Gilbert 2003; Roy and Gilbert 2006; Coulombe-Huntington and Majewski 2007a). For example, in mammals there seems to be a slow but steady rate of intron loss. Interestingly, most of the intron losses have occurred in highly expressed genes and were biased towards the 3’ end of the genes. Moreover, because small introns were found to be preferentially lost and all intron losses were precise deletions, it was concluded that the losses were likely mechanism was reverse transcription of the mature mRNA followed by homologous recombination between an intronless cDNA and the genomic version of the gene, also known as retroprocessing (Coulombe-Huntington and Majewski 2007a).

Another study on intron evolutionary dynamics was carried out using whole genome alignments from 11 Drosophila species. In this study, the authors used the currently known intron positions in *Drosophila melanogaster* to map and determine the number of gain and losses in the other Drosophila species. Not surprisingly, they found intron losses to be more prevalent and that 80% of the losses were precise deletions (consistent with the retroprocessing model). Furthermore, most of the observed losses were biased towards the 3' end of slowly evolving genes, and parallel intron losses were also observed. They concluded by saying that the overall evidence supported retroprocessing. Although not discussed in the paper, it is implied that 20% of the intron losses were imprecise deletions. These types of deletion mutations either leave behind part of the intron sequence or delete part of the exon resulting in a messy intron loss (Coulombe-Huntington and Majewski 2007b; Hepburn, Schmidt and Mower 2012).

A study on intron dynamics in fungi found that intron gains, rather than losses, were more prevalent. Furthermore, the observed losses were not skewed towards the 3' end of the gene but rather towards the center of the gene (Nielsen et al. 2004). Nielsen and co-workers found that the intron losses were inconsistent with the previous model of homologous recombination of a 3' reversed transcribed mRNA, and instead they suggested that there might be selective pressure to keep the introns located near the 5' and 3' end of the genes.

Overall, the literature seems to come to the same conclusion: there are two main mechanisms of intron loss: random genomic deletion and retroprocessing. I would like to clarify that retroprocessing encompasses two mechanisms by itself. In other words, retroprocessing is the reverse transcription of a fully spliced mRNA that undergoes homologous recombination with the native gene, deleting one or more introns. The

alternative is “retrotransposition”, in which the intronless allele arises through the same process, but retrotransposes to an ectopic location. There is a third mechanism of intron loss that is hardly ever discussed in the literature: exonization. In simple terms, the intron becomes part of the exon, occasionally giving rise to beneficial functions to the gene. Exonization appears to be more common in mammals, where introns are often short (Wang et al. 2005; Sorek 2007). For larger introns, however, exonization is unlikely to occur because the retention of an intron would generally have deleterious consequences on the gene by introducing frame shifts and premature stop codons. In general, plant group II introns are extremely large, so exonization does not seem a viable mechanism of intron loss in plant mitochondria.

We can conclude by stating that plant mitochondria genomes are an ideal system to study the mechanisms of intron loss. RNA editing serves as an additional powerful marker which ultimately would help us elucidate the mechanisms of intron loss from the mitochondrial *cox2* gene of plants.

REFERENCES

Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. 2011. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. The Plant Cell Online 23:2499-2513.

Alverson AJ, Wei XX, Rice DW, Stern DB, Barry K, Palmer JD. 2010. Insights into the evolution of mitochondrial genome size from complete sequences of Citrullus lanatus and Cucurbita pepo (Cucurbitaceae). Molecular biology and evolution 27:1436-1448.

Barkman TJ, McNeal JR, Lim SH, Coat G, Croom HB, Young ND, W dePamphilis C. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. BMC evolutionary biology 7:248.

Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants.

Bonen L. 2008. Cis-and trans-splicing of group II introns in plant mitochondria. Mitochondrion 8:26-34.

Bonen L. 2011. RNA Splicing in Plant Mitochondria. Plant Mitochondria:131-155.

Bonen L, Vogel J. 2001. The ins and outs of group II introns. TRENDS in Genetics 17:322-331.

Börner GV, Mörl M, Wissinger B, Brennicke A, Schmelzer C. 1995. RNA editing of a group II intron in Oenothera as a prerequisite for splicing. Molecular and General Genetics MGG 246:739-744.

Bruhs A, Kempken F. 2011. RNA Editing in Higher Plant Mitochondria. Plant Mitochondria:157-175.

Buchanan BB, Gruissem W, Jones RL. 2000. Biochemistry and molecular biology of plants: American Society of Plant Physiologists.

Burger G, Gray MW, Franz Lang B. 2003. Mitochondrial genomes: anything goes. TRENDS in Genetics 19:709-716.

Cho Y, Qiu YL, Kuhlman P, Palmer JD. 1998. Explosive invasion of plant mitochondria by a group I intron. Proceedings of the National Academy of Sciences 95:14244.

Coulombe-Huntington J, Majewski J. 2007a. Characterization of intron loss events in mammals. Genome research 17:23-32.

Coulombe-Huntington J, Majewski J. 2007b. Intron loss and gain in Drosophila. Molecular biology and evolution 24:2842-2850.

Gray M, Covello P. 1993. RNA editing in plant mitochondria and chloroplasts. The FASEB journal 7:64-71.

Hepburn NJ, Schmidt DW, Mower JP. 2012. Loss of two introns from the Magnolia tripetala mitochondrial cox2 gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. Molecular biology and evolution.

Janska H, Mackenzie S. 1993. Unusual mitochondrial genome organization in cytoplasmic male sterile common bean and the nature of cytoplasmic reversion to fertility. Genetics 135:869-879.

Janska H, Sarria R, Woloszynska M, Arrieta-Montiel M, Mackenzie SA. 1998. Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. The Plant Cell Online 10:1163-1180.

Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics 9:605-618.

Knoop V, Volkmar U, Hecht J, Grewe F. 2011. Mitochondrial genome evolution in the plant lineage. Plant Mitochondria:3-29.

Lambowitz AM, Caprara MG, Zimmerly S, Perlman PS. 1999. Group I and group II ribozymes as RNPs: clues to the past and guides to the future. COLD SPRING HARBOR MONOGRAPH SERIES 37:451-486.

Lambowitz AM, Zimmerly S. 2004. Mobile group II introns. Annu. Rev. Genet. 38:1-35.

Lambowitz AM, Zimmerly S. 2011. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harbor Perspectives in Biology 3.

Lang BF, Gray MW, Burger G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. Annual review of genetics 33:351-397.

Li-Pook-Than J, Carrillo C, Niknejad N, Calixte S, Crosthwait J, Bonen L. 2007. Relationship between RNA splicing and exon editing near intron junctions in wheat mitochondria. Physiologia Plantarum 129:23-33.

Li W, Tucker AE, Sung W, Thomas WK, Lynch M. 2009. Extensive, recent intron gains in Daphnia populations. Science 326:1260-1262.

Logan DC. 2006. Plant mitochondrial dynamics. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research 1763:430-441.

Lonsdale D, Brears T, Hodge T, Melville SE, Rottmann W. 1988. The plant mitochondrial genome: homologous recombination as a mechanism for generating heterogeneity. Philosophical Transactions of the Royal Society of London. B, Biological Sciences 319:149-163.

Mackenzie S, McIntosh L. 1999. Higher plant mitochondria. The Plant Cell Online 11:571-586.

Majewski J, Ott J. 2002. Distribution and characterization of regulatory elements in the human genome. Genome research 12:1827-1836.

Maréchal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. New Phytologist 186:299-317.

Mourier T, Jeffares DC. 2003. Eukaryotic intron loss. Science 300:1393-1393.

Mower JP. 2008. Modeling sites of RNA editing as a fifth nucleotide state reveals progressive loss of edited sites from angiosperm mitochondria. Molecular biology and evolution 25:52-61.

Mower JP, Sloan DB, Alverson AJ. 2012. Plant mitochondrial genome diversity: the genomics revolution. Plant Genome Diversity Volume 1:123-144.

Mueller JE, Smith D, Belfort M. 1996. Exon coconversion biases accompanying intron homing: battle of the nucleases. Genes & Development 10:2158-2166.

Nielsen CB, Friedman B, Birren B, Burge CB, Galagan JE. 2004. Patterns of intron gain and loss in fungi. PLoS biology 2:e422.

Richardson AO, Palmer JD. 2007. Horizontal gene transfer in plants. Journal of experimental botany 58:1-9.

Roy SW, Fedorov A, Gilbert W. 2003. Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. Proceedings of the National Academy of Sciences of the United States of America 100:7158.

Roy SW, Gilbert W. 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. Nature Reviews Genetics 7:211-221.

Sanchez-Puerta MV, Abbona CC, Zhuo S, Tepe EJ, Bohs L, Olmstead RG, Palmer JD. 2011. Multiple recent horizontal transfers of the cox1 intron in Solanaceae and extended co-conversion of flanking exons. BMC evolutionary biology 11:277.

Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD. 2008. Frequent, phylogenetically local horizontal transfer of the cox1 group I intron in flowering plant mitochondria. Molecular biology and evolution 25:1762-1777.

Shikanai T. 2006. RNA editing in plant organelles: machinery, physiological function and evolution. Cellular and molecular life sciences 63:698-708.

Simon DM, Kelchner SA, Zimmerly S. 2009. A broadscale phylogenetic analysis of group II intron RNAs and intron-encoded reverse transcriptases. Molecular biology and evolution 26:2795-2808.

Sloan D, Alverson A, Štorchová H, Palmer J, Taylor D. 2010. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm Silene latifolia. BMC evolutionary biology 10:274.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012. Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. PLoS biology 10:e1001241.

Small I, Suffolk R, Leaver C. 1989. Evolution of plant mitochondrial genomes via substoichiometric intermediates. Cell 58:69-76.

Sorek R. 2007. The birth of new exons: mechanisms and evolutionary consequences. Rna 13:1603-1608.

Vaughn JC, Mason MT, Sper-Whitis GL, Kuhlman P, Palmer JD. 1995. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric coxI gene of Peperomia. Journal of molecular evolution 41:563-572.

Wang W, Zheng H, Yang S, Yu H, Li J, Jiang H, Su J, Yang L, Zhang J, McDermott J. 2005. Origin and evolution of new exons in rodents. Genome research 15:1258-1264.

Woloszynska M, Bocer T, Mackiewicz P, Janska H. 2004. A fragment of chloroplast DNA was transferred horizontally, probably from non-eudicots, to mitochondrial genome of Phaseolus. Plant molecular biology 56:811-820.

CHAPTER 2

Loss of two introns from the *Magnolia tripetala* mitochondrial *cox2* gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss

Nancy J. Hepburn[1,2], Derek W. Schmidt[1,3], and Jeffrey P. Mower[1,2]

[1] Center for Plant Science Innovation, [2] Department of Agronomy and Horticulture, and [3] School of Biological Sciences, University of Nebraska, Lincoln, NE 68588, USA

This chapter was published exactly as presented in Molecular Biology and Evolution

Hepburn, N.J., Schmidt, D.W. and Mower, J.P. (2012) Loss of two introns from the Magnolia tripetala mitochondrial *cox2* gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. *Molecular Biology and Evolution*, inpress.

ABSTRACT

Intron loss is often thought to occur through retroprocessing, the reverse transcription and genomic integration of a spliced transcript. In plant mitochondria, several unambiguous examples of retroprocessing are supported by the parallel loss of an intron and numerous adjacent RNA edit sites, but in most cases the evidence for intron loss via retroprocessing is weak or lacking entirely. To evaluate mechanisms of intron loss, we designed a PCR-based assay to detect recent intron losses from the mitochondrial *cox2* gene within genus *Magnolia*, which was previously suggested to have variability in *cox2* intron content. Our assay showed that all 22 examined species have a *cox2* gene with two introns. However, one species, *M. tripetala*, contains an additional *cox2* gene that lacks both introns. Quantitative PCR showed that both *M. tripetala cox2* genes are present in the mitochondrial genome. Although the intronless gene has lost several ancestral RNA edit sites, their distribution is inconsistent with retroprocessing models. Instead, phylogenetic and gene conversion analyses indicate that the intronless gene was horizontally acquired from a eudicot and then underwent gene conversion with the native intron-containing gene. Models are presented to summarize the roles of horizontal gene transfer and gene conversion as a novel mechanism of intron loss.

INTRODUCTION

Intron content is highly variable among eukaryotic genes and genomes (reviewed in Belshaw, Bensasson 2006; Lang, Laforest, Burger 2007; Schmitz-Linneweber, Barkan 2007), indicating frequent gain and loss of introns over evolutionary time. Plant mitochondrial genomes offer an intriguing system in which to study intron evolutionary dynamics. To date, there are >50 complete genomes available, with representatives from all major land plant and green algal groups, and nearly all of them carry group I and/or group II introns (reviewed in Bonen 2011; Mower, Sloan, Alverson in press). Intron content varies dramatically among species, from 0 in some algae to 37 in the spikemoss *Selaginella moellendorffii* (Robbens et al. 2007; Hecht, Grewe, Knoop 2011). Among angiosperms, intron content appears to have mostly stabilized. Exceptions include a few rare lineage-specific intron losses from some genomes (Kubo et al. 2000; Sugiyama et al. 2005; Sloan et al. 2010a; Mower et al. in press) and the horizontal acquisition of a group I intron by various angiosperms (Cho et al. 1998; Sanchez-Puerta et al. 2008). The widely studied introns in the *cox2* gene also show extensive variation among species (Hiesel, Brennicke 1983; De Benedetto et al. 1992; Rabbi, Wilson 1993; Albrizio et al. 1994; Qiu et al. 1998; Joly, Brouillet, Bruneau 2001; Kudla et al. 2002), suggesting a dynamic evolutionary history involving frequent losses from this gene.

Given the prevalence of intron gain and loss throughout eukaryotic evolution, the mechanisms that govern these processes have been extensively studied. Because group I and II introns are mobile elements (at least historically), their mechanisms of spread are well known (Lambowitz, Belfort 1993; Lambowitz, Zimmerly 2011). Possible mechanisms of intron loss include genomic deletion, exonization, and retroprocessing, each of which is expected to have different effects on gene structure. In exonization, the intron sequence is no longer spliced out of the transcript; instead, it is retained in the

mature transcript and translated, which introduces novel amino acids into the produced protein (Parma et al. 1987; Wang et al. 2005). Intron loss by direct genomic deletion will often imprecisely remove the intron (Llopart et al. 2002; Coulombe-Huntington, Majewski 2007), such that adjacent exonic sequences may also be deleted (resulting in the loss of encoded amino acids) or small intron fragments may be retained (which become exonized). In retroprocessing, intron loss occurs through the genomic integration of a cDNA intermediate created by reverse transcription of an intron-spliced transcript (Fink 1987; Derr, Strathern 1993). This mechanism is expected to precisely remove introns, with a bias towards 3' intron removal (Mourier, Jeffares 2003; Roy, Gilbert 2005). In addition, because plant mitochondrial transcripts undergo extensive cytosine-to-uracil (C-to-U) RNA editing (and U-to-C editing in some lineages), retroprocessing will produce intronless genes that also contain the edited nucleotide states, which effectively eliminates the edit sites from the retroprocessed gene. The parallel loss of RNA editing provides a strong additional marker of intron loss via retroprocessing, making plant mitochondria an ideal system in which to test for this mechanism.

The few examined cases of intron loss from plant mitochondrial genes indicate, or at least suggest, a role for retroprocessing. Three intron losses from the lycophyte *Isoetes engelmannii* (Grewe et al. 2011) and two from gymnosperms (Ran, Gao, Wang 2010) provide the strongest evidence for retroprocessing, because numerous edit sites that flanked the introns were lost in parallel. Several additional examples of intron loss via retroprocessing have been reported, but in these cases only a few edit sites were lost (Geiss, Abbas, Makaroff 1994; Itchoda et al. 2002; Sloan et al. 2010b). For loss via genomic deletion there is only a single report, in which a *Petunia* cytoplasmic male-sterility gene containing an intronless fragment of *cox2* co-exists in the genome with a full-length, intron-containing *cox2* gene (Pruitt, Hanson 1989). In this *cox2* fragment,

exon sequences on both sides of the missing intron are also absent, suggesting intron loss via imprecise genomic deletion. There is no evidence for intron loss via exonization in plant mitochondrial genomes, which is not surprising given the large size of these introns. For many cases of plant mitochondrial intron loss, the mechanism of loss was not reported (e.g., Qiu et al. 1998; Joly, Brouillet, Bruneau 2001; Kudla et al. 2002; Sugiyama et al. 2005; Mower et al. in press), which suggests that the authors did not look or that the evidence was lacking.

The study of recent intron losses could help determine the underlying mechanism of intron loss from plant mitochondrial genes. In our initial survey of the literature and the sequences in GenBank for examples of plant mitochondrial intron loss, the *cox2* gene from genus *Magnolia* stood out because it exhibited intron variability among species, suggesting one or more recent cases of intron loss. These preliminary findings prompted us to do an expanded survey of *cox2* intron content in *Magnolia* and related magnoliids.

MATERIALS AND METHODS

Plant Materials

Leaf material from four *Magnolia* species (*M. acuminata*, *M. kobus*, *M. stellata*, *M. tripetala*) and most other genera was collected from the Earl G. Maxwell Arboretum or from the living collection at the Beadle Center Greenhouse (University of Nebraska-Lincoln). Leaf material from an additional 18 *Magnolia* species was provided by Richard B. Figlar (JC Raulston Arboretum, North Carolina). Saplings of *M. pyramidata* and a second individual of *M. tripetala* were obtained from Nearly Native Nursery (Fayetteville, GA). Source of material and voucher information for all plant species used in this study are provided in Supplementary Table 1.

Molecular biology techniques

Total genomic DNA and RNA was isolated from fresh leaf tissue using DNeasy Plant Mini or Maxi Kits and RNeasy Plant Mini Kits (QIAGEN). Total RNAs were treated with DNase I (Fermentas) to eliminate DNA contamination according to the manufacturer's instructions. First-strand cDNA was synthesized from random hexamers using the RevertAid First Strand cDNA Synthesis Kit (Fermentas). To ensure that there was no DNA contamination, a control sample was also prepared without reverse transcriptase. This control sample was evaluated alongside the first strand cDNA preparation in downstream analyses.

Polymerase chain reaction (PCR) assays were performed using degenerate primers (Supplementary Table 2) and GoTaq Flexi DNA Polymerase with supplied reagents (Promega). Each reaction was amplified in a PTC-0220G or C1000 thermal cycler (Bio-Rad). PCR reactions for all intron-containing gene assays included an initial denaturation step (94°C for 3 min), 35 cycles of denaturation (94°C for 45 sec), annealing (54°C for 1 min), and elongation (72°C for 3 min), and a final elongation step (72°C for 10 min). For the intronless gene assay, the annealing and extension times were reduced to 30 sec to prevent amplification of large, intron-containing products.

Reverse-transcription PCR (RT-PCR) assays were performed similarly to the PCR assays using first strand cDNA as template and degenerate primers (Supplementary Table 2). The RT-PCR program included an initial denaturation step (94°C for 3 min), 35 cycles of denaturation (94°C for 30 sec), annealing (50°C for 45 sec), and elongation (72°C for 2 min), and a final elongation step (72°C for 5 min).

Quantitative PCR (qPCR) assays were performed using the iCycler iQ system (Bio-Rad). Primers (Supplementary Table 2) were designed using the Integrated DNA Technologies

RealTime PCR tool (http://www.idtdna.com/scitools/) with default parameters. Each 20 µl qPCR reaction contained 10 µl of iQ SYBR Green Supermix (Bio-Rad), 250 nM of each primer, and 20 ng of DNA. Triplicate reactions were carried out with an initial denaturation step (95°C for 3 min) followed by 40 cycles of denaturation (95°C for 10 sec) and annealing/elongation (60°C for 45 sec). To check for the presence of multiple products, a melt-curve analysis was performed at the end of the run for each reaction (60–100°C at 0.5°C increments for 10 sec per step).

Sequencing and Sequence Analysis

PCR and RT-PCR amplicons were purified and directly sequenced on both strands at the High-Throughput Genomics Unit (University of Washington, Seattle, USA). Sequences were assembled with CodonCode Aligner version 3.5 (CodonCode Corporation). RNA edit sites were experimentally determined by amplifying and sequencing cDNA products using primers described in the text and then comparing these sequences to the DNA sequences. All newly generated sequences were deposited in GenBank under accession numbers JQ317131–JQ317153, and additional sequences were obtained from GenBank (Supplementary Table 3).

Exon sequences were aligned using MUSCLE version 3.7 (Edgar 2004) and manually refined using BioEdit (Hall 1999). Poor quality regions were removed by GBlocks version 0.91b (Castresana 2000) using relaxed parameters (b2=half+1, b4=5, b5=half). Phylogenetic analysis was performed using PhyML 3.0 (Guindon et al. 2010). The maximum likelihood analysis used the general time-reversible substitution model and the subtree pruning and regrafting tree-search method. The proportion of invariable sites and the shape of the gamma rate distribution with four categories were estimated during the analysis. Data sets were analyzed ten times using randomized starting tress. Tree support was evaluated from 1000 bootstrap replicates. Trees were rooted on the basal

angiosperms *Amborella trichopoda* and *Schisandra chinensis*. For the RNA phylogenetic analysis, edit sites were converted from C to T in the data set. Most edit sites were experimentally determined, either in this study or from annotations in GenBank accession files. These RNA sequences were used were predict edit sites for the remaining species using PREP-Aln (Mower 2009) with a cutoff value set to 0.5.

Gene chimerism was tested using the Comp3Seq program (http://www.indiana.edu/~orgconv/) in the OrgConv package (Hao 2010). For this analysis, the *M. tripetala* intronless gene sequence was compared to consensus sequences from Magnoliales and from core eudicots.

Species	Intron content [a]		Source of evidence
	cox2-i1	*cox2*-i2	
Laurales			
Calycanthus fertilis	−	?	PCR survey (Joly, Brouillet, Bruneau 2001)
Laurus nobilis	+	?	Sequencing (GenBank AY832094, AY832104)
Magnoliales			
Liriodendron tulipifera	+	+	Sequencing (GenBank AY832090, AY832101; AO Richardson, JD Palmer, unpublished data), PCR survey (Joly, Brouillet, Bruneau 2001)
Magnolia grandiflora	+	?	Sequencing (GenBank X78418), Southern blot (De Benedetto et al. 1992)
Magnolia liliiflora	+	?	PCR survey (Joly, Brouillet, Bruneau 2001)
Magnolia pyramidata	−	?	Sequencing (GenBank U42696)
Magnolia soulangeana	+	?	Southern blot (De Benedetto et al. 1992)
Magnolia tripetala	−	?	Sequencing (GenBank U43055)
Piperales			
Asarum sp.	+	?	Sequencing (GenBank AY832096, AY832105)
Peperomia obtusiolia	+	+	Southern blot (Qiu et al. 1998)
Piper betle	+	+	Sequencing (GenBank AY832091, AY832102), Southern blot (Qiu et al. 1998), PCR survey (Joly, Brouillet, Bruneau 2001)
Saururus chinensis	+	+	Southern blot (Qiu et al. 1998)

[a] intron present (+), absent (−), status unknown (?)

Table1. Previously reported *cox2* intron distribution in magnoliids

RESULTS

Magnolia tripetala *contains two mitochondrial versions of* cox2 *that differ in intron content*

To evaluate the extent and timing of *cox2* intron loss in *Magnolia*, we designed a PCR screen that specifically targets intron-containing and intronless versions of *cox2* (Fig. 1A). Using the primers designed to amplify *cox2* genes with one or more introns, we found that two introns were present in all 22 examined *Magnolia* species (Fig. 1B; Supp. Fig. 1A and 1B), including *M. pyramidata* and *M. tripetala*, which were expected to lack at least the first intron based on available GenBank accessions (Table 1). The two introns have the nomenclatural designation *cox2i373* and *cox2i691* based on their nucleotide positioning relative to the *cox2* gene from the liverwort *Marchantia polymorpha* (Dombrovska, Qiu 2004), but we will hereafter referred to them as i1 and i2 for simplicity.

Interestingly, when we used the primers and PCR protocol specifically designed to amplify an intronless *cox2* gene, we obtained a product from *M. tripetala* that was comparable in size to the product from *Geranium sanguineum*, whose *cox2* gene is intronless (Fig. 1C). We repeated the intronless PCR assay using DNA from an independently acquired *M. tripetala* individual, which again produced a putative intronless *cox2* product (Supp. Fig. 1C). Sequencing of these short PCR products from the two *M. tripetala* individuals confirmed that they are indeed intronless copies of *cox2*; furthermore, their sequences are identical to one another and >99% similar to the intronless *M. tripetala* sequence available in GenBank. Thus, there appear to be two loci for *cox2* in *M. tripetala*, one with two introns and another that lacks both introns. For *M.*

pyramidata, however, our screen for intronless *cox2* genes failed to recover a product from our individual (Fig. 1C), in contrast to expectations based on the intronless sequence from this species available in GenBank.

Using a previously developed qPCR strategy to determine the cellular genomic location of a gene (Mower et al. 2010), we found that the intron-containing and intronless *cox2* genes in *M. tripetala* are both present in the mitochondrial genome (Fig. 1D). For this assay, we designed two sets of qPCR primers that specifically target the intron-containing copy by amplifying from the intron sequences, and two other sets of primers that specifically target the intronless copy by amplifying across the exon junctions (Fig. 1A). All four *cox2* products showed similar cycle numbers to one another and to known mitochondrial genes (*atp1*, *rpl10*), indicating that the two *M. tripetala cox2* genes are located in the mitochondrial genome (Fig. 1D). As expected, known plastid genes (*matK, ndhF*) appeared first in the assay, mitochondrial genes appeared next, and nuclear genes (*Gai*1, *Lfy*, *Phy*A) appeared last, which is consistent with copy number expectations for the plastid (highest), mitochondrial (intermediate), and nuclear (lowest) genomes in a plant cell (Lamppa, Bendich 1984; Draper, Hays 2000).

FIG 1

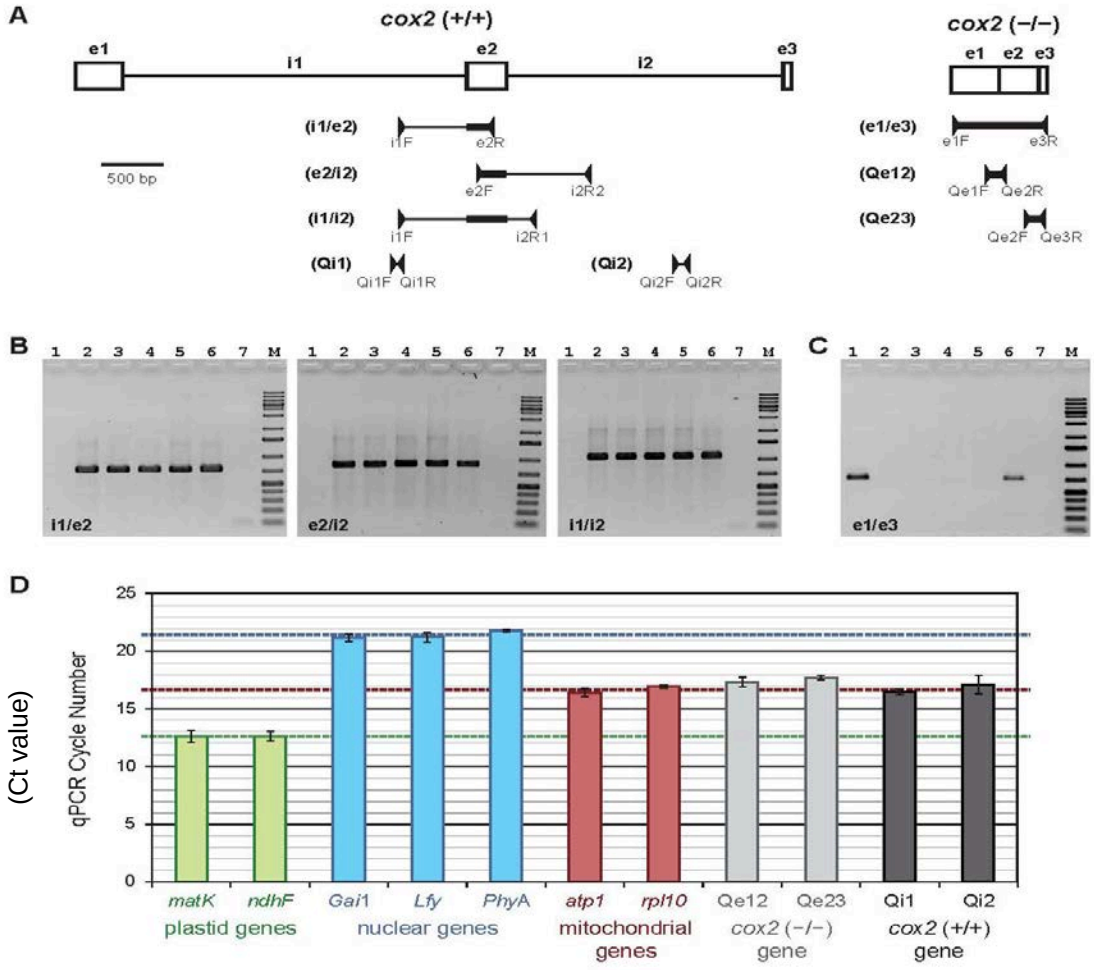


Figure 1. Intron status of the *cox2* gene in Magnoliales. A) Design of the PCR and qPCR assays to specifically target the intron-containing (+/+) or intronless (−/−) *cox2* gene. Exon and introns were drawn to scale of the *Liriodendron cox2* gene. B and C) Assay results using primers designed to amplify a *cox2* gene with intron 1 (i1/e2), intron 2 (e2/i2), both introns (i1/i2), or no introns (e1/e3). Lane 1: *Geranium sanguineum*, 2: *Liriodendron tulipifera*, 3: *Magnolia acuminata*, 4: *M. kobus*, 5: *M. pyramidata*, 6: *M. tripetala*, 7: negative control, M: GeneRuler 1 kb plus DNA ladder (Fermentas). D) Cellular genomic location of the two *M. tripetala cox2* genes. qPCR assay results using primers designed to amplify the plastid *matK* and *ndhF* genes (in green); the nuclear *Gai*1, *Lfy*, and *Phy*A genes (in blue); the mitochondrial *atp1* and *rpl10* genes (in red); the exon 1/2 junction (Qe12) and the exon 2/3 junction (Qe23) of the intronless *cox2* gene (in light gray); and intron 1 (Qi1) and intron 2 (Qi2) from the intron-containing *cox2* gene (in dark gray). Horizontal dashed lines indicate the mean qPCR values for the known plastid (green), nuclear (blue), or mitochondrial genes (red).

Unexpressed intronless cox2 *gene*

To examine the relative expression levels of the two *M. tripetala* genes, we amplified and directly sequenced the e1F/e3R RT-PCR product (Fig. 2). If both genes are expressed, the directly sequenced cDNA should exhibit sequence polymorphism at all four sites that differ between the intron-containing and intronless genes (ignoring C and T differences between the genes at sites of RNA editing; see next section). However, only one peak was visible in the cDNA electropherogram at each of these four sites, and each peak was identical to the intron-containing gene sequence, indicating that the cDNA was derived from the intron-containing gene only. Thus, there are few if any transcripts produced by the intronless gene in the mitochondrial RNA pool, at least in mature leaf tissue from which the RNA was extracted.

FIG 2

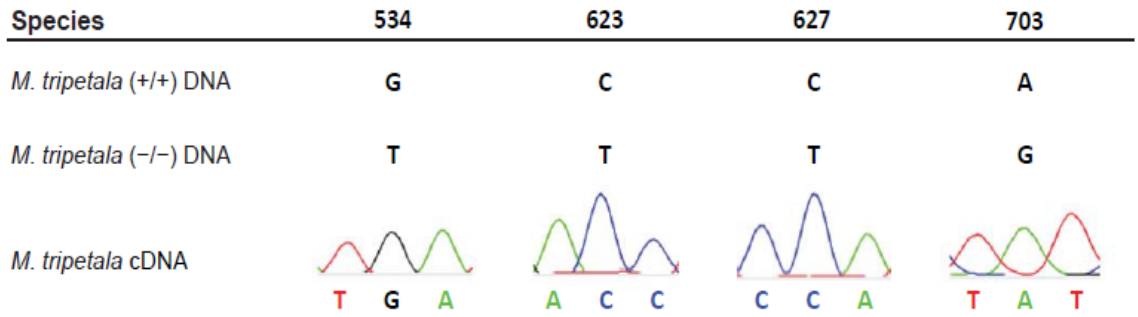


Figure 2. Evaluation of transcription of the two *M. tripetala cox2* genes. The electropherogram of the cDNA sequence amplified using primers e1F and e3R is shown at sites that differ between the intron-containing (+/+) and intronless (−/−) *cox2* genes. Numbers correspond to positions in the *Liriodendron* coding sequence.

Patterns of edit site loss from the intronless* cox2 *gene are inconsistent with retroprocessing

The presence of an intron-containing and an intronless version of the same gene in the same genome suggested that the intronless gene may have arisen through retroprocessing of the intron-containing gene. We designed a RT-PCR assay to identify sites of RNA editing in the intron-containing copy at various stages of splicing (Fig. 3A) and then evaluated the intronless copy at these sites to check for evidence of retroprocessing (Fig. 3B and 3C).

We identified 17 sites at which C is edited to U in transcripts from the intron-containing cox2 genes from several *Magnolia* species, and 18 in other Magnoliales (Fig. 3B). If retroprocessing were to occur from a fully spliced and edited transcript from the intron-containing gene, we would expect that, along with the loss of both introns, all of the U's produced by editing would be seen as T's in the retroprocessed gene sequence. However, in the intronless *cox2* gene, 13 of the 17 *Magnolia* edit positions have a genome-encoded C instead of a T, arguing against a retroprocessing model affecting the entire mature transcript (Fig. 3B).

Another model of retroprocessing involves a localized event that affects only part of the gene, thereby removing the intron and nearby edit sites but not any distant edit sites. In many reported cases of retroprocessing, the pattern of edit site loss is consistent with such a localized retroprocessing event (Geiss, Abbas, Makaroff 1994; Itchoda et al. 2002; Sloan et al. 2010b; Grewe et al. 2011). Again, however, the observed pattern of edit site loss in the intronless gene is inconsistent with this localized retroprocessing model (Fig. 3B). None of the closest edit sites on either side of either intron were lost, not even at positions 379 or 698, edit sites which are only a few nucleotides away from the site of intron loss.

A third model of retroprocessing involves the reverse transcription of a spliced but mostly unedited transcript, resulting in the sporadic loss of some but not all edit sites from the retroprocessed gene. If this were the case, we would expect RNA editing to be slow relative to splicing in the transcript, so that there would be an appreciable number of spliced but incompletely edited transcripts to serve as a retroprocessing template. To test this possibility, we compared the completeness of editing in transcripts from the intron-containing gene at various stages of splicing (Fig. 3C). By examining the relative height of T and C electropherogram peaks from the directly sequenced transcripts, it is clear that most editing activity occurs prior to or in parallel with intron splicing, with a possible exception at position 460, which is a silent editing event. In unspliced transcripts, all edit sites (except 460) show significant editing activity (25–74%). By the time that one of the introns is spliced, editing activity at most sites is mostly complete (75–100%), although there appears to be a delay in editing at positions 676 and 695 when intron 2 is still present (suggesting that these edit sites are important for or dependent upon intron 2 splicing). In fully spliced transcripts, all sites except 460 are essentially fully edited (94–100%). Given the completeness of editing after splicing, it is extremely unlikely that the *M. tripetala* intronless *cox2* gene was generated via retroprocessing of a spliced but mostly unedited transcript from the intron-containing gene.

FIG 3

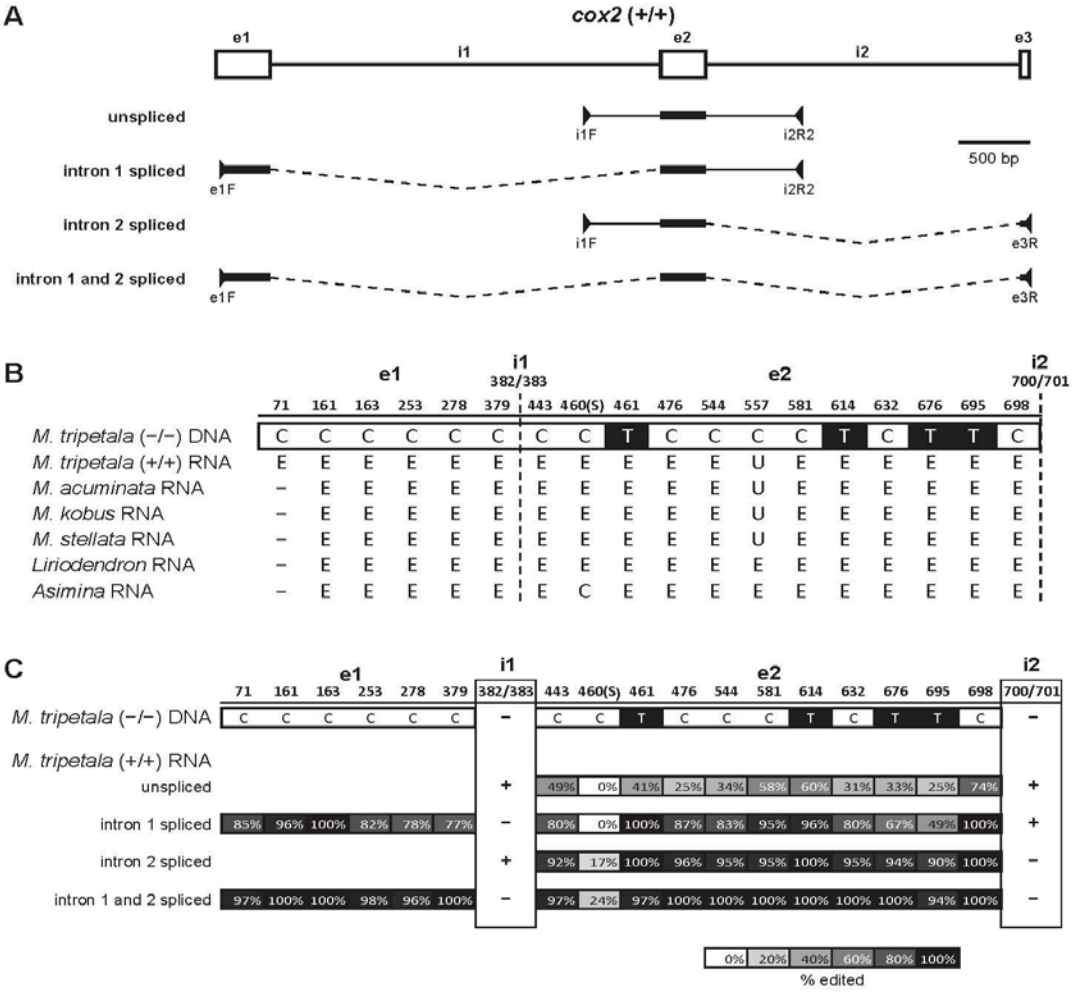


Figure 3. Evaluation of RNA editing and retroprocessing. A) Design of the RT-PCR to specifically target fully-spliced or immature transcripts. Exon and introns were drawn to scale of the *Liriodendron cox2* gene. B) Nucleotide state of the *M. tripetala* intronless *cox2* gene at sites of RNA editing in other Magnoliales. RNA edit sites (denoted as 'E') were experimentally determined; their nucleotide positions correspond to the full-length *cox2* coding sequence from *Liriodendron*. The silent edit site at position 460 is marked with an S. Unedited cytosines ('C') and genome-encoded thymines producing uracils ('U') in the transcript are also shown where present. At each edit site, genome-encoded thymines ('T') in the *M. tripetala* intronless *cox2* DNA sequence are highlighted in black. C) Evaluation of editing completeness in transcripts at various stages of splicing. Edited frequency was calculated by comparing the peak heights for T and C in the electropherogram, approximating the frequency of transcripts that are edited at each site in the RNA pool. Sites are highlighted according to their edited frequency using a gradient from white (fully unedited) to black (fully edited).

The intronless* cox2 *gene has a chimeric structure implicating horizontal gene transfer and gene conversion

Close inspection of the *M. tripetala* intronless *cox2* sequence indicated a non-random distribution of sequence divergence (Fig. 4A). The beginning (nucleotides 57–460 relative to *Liriodendron tulipifera*) and end (696–741) of this DNA sequence are nearly identical (>99%) to the intron-containing *cox2* sequences from other *Magnolia* species, whereas the middle (461–695) is more similar to sequences from core eudicots than from other *Magnolia* species. Interestingly, all four of the ancestral *Magnolia* edit sites that are genome-encoded T's in the *M. tripetala* intronless gene are located in the eudicot-like part of the sequence, and most core eudicots also predominantly contain T at these positions. Using a statistical test to evaluate whether the *M. tripetala* intronless gene is a chimeric gene with *Magnolia*-like and eudicot-like components, we found strong support ($P = 2x10^{-7}$) for a chimeric structure with recombination breakpoints at positions 461 and 695. The chimeric nature of the *M. tripetala* intronless gene suggested that it may be derived from a combination of horizontal transfer of a eudicot *cox2* gene and gene conversion with the native gene.

Although the chimeric gene test defined the minimal size of the chimeric region in the *M. tripetala* intronless sequence to be between nucleotides 461 and 695, the limited divergence between the consensus sequences from Magnoliales and core eudicots makes it impossible to identify the exact recombination breakpoints (Fig. 4B). The most likely place for the upstream breakpoint is between nucleotides 251 and 460, because the *M. tripetala* intronless sequence is clearly *Magnolia*-like before nucleotide 251 and clearly eudicot-like after 460. This stretch of DNA spans the intron 1 position, which means that the precise upstream breakpoint could have occurred within exon 1 or exon 2. The downstream breakpoint is similarly uncertain. Between nucleotides 696 and 720,

there is one site that weakly supports an association with Magnoliales, another site that weakly supports an association with core eudicots, but no sites that provide strong support for either association. Because the intron 2 position is within this region, the downstream recombination breakpoint could either be near the end of exon 2 or the beginning of exon 3.

To determine the evolutionary origins of the different parts of the *M. tripetala* intronless *cox2* gene, we split the gene into its two components (*Magnolia*-like and eudicot-like) and then performed a phylogenetic analysis with exon sequences from representatives of the Magnoliales and other flowering plants (Fig. 4C). In general, the tree showed overall relationships that largely agree with currently accepted taxonomy (http://www.mobot.org/mobot/research/apweb/), recovering monophyletic groups with weak to strong bootstrap support for the major angiosperm clades including magnoliids (62%), monocots (95%), and eudicots (72%). As expected, the *Magnolia*-like component of the *M. tripetala* sequence grouped with other magnoliids with good support (86%), and it specifically clustered with the intron-containing *Magnolia* sequences, albeit with no support (<50%). In contrast, the 235 bp eudicot-like portion of the *M. tripetala* intronless gene clustered away from other magnoliids and grouped instead within eudicots (72%) and more specifically within core eudicots (56%). Interestingly, this sequence was found in a clade of rosids that includes *Pisum* and *Oenothera*, whose *cox2* genes lack both introns.

We also examined the phylogenetic placement of the *M. pyramidata* intronless sequence from GenBank (Fig. 4C), which does not have an obvious chimeric structure ($P > 0.05$). This sequence grouped within eudicots rather than within magnoliids, but it did not show any clear affinity with the eudicot-like part of the *M. tripetala* intronless gene. This placement is consistent with another horizontal transfer event from some eudicot donor

into *M. pyramidata*. However, because we were unable to recover the same intronless sequence from our own *M. pyramidata* individual, we cannot exclude the possibility that the *M. pyramidata* intronless sequence in GenBank was amplified from a misidentified or contaminated DNA sample.

Because RNA edit sites can cause artifacts during phylogenetic analysis, we performed an additional analysis using an RNA data set, which effectively removes any RNA editing effects (Supp. Fig. 2). This reanalysis did not qualitatively affect the previous phylogenetic results, although bootstrap support values were reduced to insignificant levels for several branches.

FIG 4

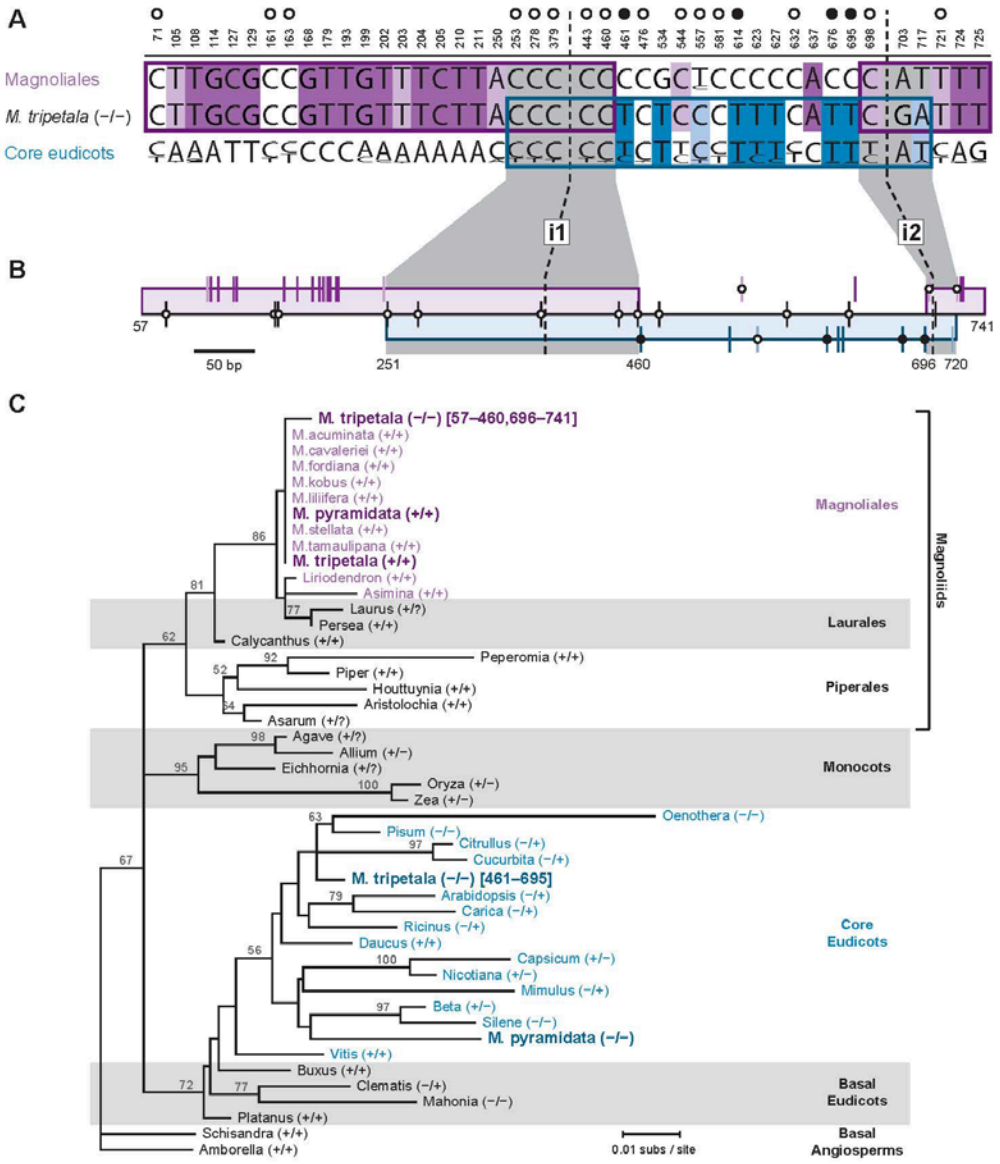


Figure 4.. Analysis of the chimeric structure of the *M. tripetala* intronless *cox2* gene. A) Shown are all *cox2* positions that are variable among the *M. tripetala* intronless sequence and the consensus sequences from Magnoliales and core eudicots. Consensus sequences are presented as sequence logos, in which the height of the nucleotide letter indicates the frequency of that nucleotide in the sequences. *M. tripetala* nucleotides are shaded in purple if they are identical to the Magnoliales consensus and in blue if they agree with the core eudicot consensus. Dark shading indicates strong support and light shading indicates weaker support because of some ambiguity from a minor frequency variant in one of the consensus sequences. Colored boxes outline the possible extremities of the chimeric pieces. Gray shading indicates the most likely recombination points that formed the chimeric sequence. Circles above the nucleotide positions indicate sites of RNA editing; filled circles indicate the positions that potentially support retroprocessing from Figure 3B. Numbers correspond to positions in the *Liriodendron* coding sequence. B) Graphical representation of the chimeric structure of the *M. tripetala* intronless *cox2* gene drawn to scale. Vertical lines indicate positions from Figure 4A that support an association with Magnoliales (top in purple), core eudicots (bottom in blue) or neither (middle in black). Circles and shading corresponds to Figure 4A. C) Phylogenetic analysis of the Magnoliales-like portion (positions 57–460 and 696–741) and the eudicot-like portion (positions 461–695) of the *M. tripetala* intronless *cox2* gene. The *cox2* intron status is given to the right of each species name. DNA sequences (exons only) were used for all species. Bootstrap values from 1000 replicates are shown when >50%. Tree was rooted on *Amborella trichopoda* and *Schisandra chinensis*. M. = *Magnolia*.

DISCUSSION

We have shown that *M. tripetala* contains two mitochondrial versions of the *cox2* gene that differ in intron content: one gene contains two introns whereas the other lacks both introns (Fig. 1). Most other magnoliids also contain both introns, indicating that the *cox2* gene probably contained two introns in the magnoliid common ancestor. The phylogenetic placement of the *M. tripetala* intron-containing gene with other magnoliid intron-containing genes confirms that this version was vertically acquired from the magnoliid ancestor.

How did the intronless version arise? The simplest explanation is that the intronless gene arose through retroprocessing of the intron-containing gene. However, the sporadic distribution of edit site loss could not be reconciled with various models of retroprocessing involving 1) the entire gene, 2) localized regions of the gene, or 3) incompletely edited transcripts (Fig. 3). An alternative explanation for intron loss is horizontal gene transfer of a eudicot intronless gene and gene conversion with the *Magnolia* intron-containing gene (Fig. 4). There is a clear and statistically significant chimeric signal in the middle of the *M. tripetala* intronless *cox2* gene, which is most similar to the homologous region in eudicot *cox2* genes. Although four of the nucleotide substitutions in this region compared to the Magnoliales consensus could be explained either by HGT or retroprocessing, there are another five substitutions (three nonsynonymous, two synonymous) that are consistent with HGT but not with retroprocessing. It is these additional sites that provided the HGT signal in the phylogenetic analysis of RNA data (which eliminates the four ambiguous HGT/retroprocessing sites). Furthermore, there are no sites that unambiguously support retroprocessing over HGT, despite the fact that 11 edit sites (71, 161, 163, 253, 278,

379, 443, 460, 476, 581, 632) would have provided such unambiguous support if they were lost via retroprocessing.

On balance, a model involving HGT and gene conversion provides a more comprehensive explanation for the chimeric gene structure of the *M. tripetala* intronless *cox2* gene. HGT is well-known to affect plant mitochondrial genomes (Richardson, Palmer 2007; Bock 2010), and several recent reports have shown that gene conversion can occur between foreign and native genes that co-reside in the plant mitochondrial genome (Bergthorsson et al. 2003; Barkman et al. 2007; Hao, Palmer 2009; Hao et al. 2010; Mower et al. 2010; Sloan et al. 2010a). We present two models to show how HGT and gene conversion can mediate intron loss from a gene (Fig. 5). In both models, the horizontally acquired gene arrives without introns, suggesting that a species lacking introns was the HGT donor. Given the small size of the chimeric region, it is unlikely that we will be able to unambiguously identify the eudicot donor species. However, it is intriguing that the chimeric region groups within rosids, which are known to have lost intron 1 early in evolution (Joly, Brouillet, Bruneau 2001). Less is known about the distribution of intron 2 among angiosperms, but it is certainly absent from a number of rosid lineages including Fabales [e.g., *Pisum* (GenBank accession AJ414385), *Glycine* (X04825)], Geraniales [*Geranium* (this study), *Pelargonium* (DQ317069)], Myrtales [*Oenothera* (X04202)], and at least some Cucurbitales [*Cucumis melo* (JF412792), *C. sativus* (HQ860792)].

This is the first report documenting HGT and gene conversion as a mechanism of intron loss. Given the generally weak evidence in support of retroprocessing in many plant mitochondrial studies of intron loss, it would be useful to re-evaluate these other cases to determine whether HGT and gene conversion may provide a better explanation for intron loss.

FIG 5

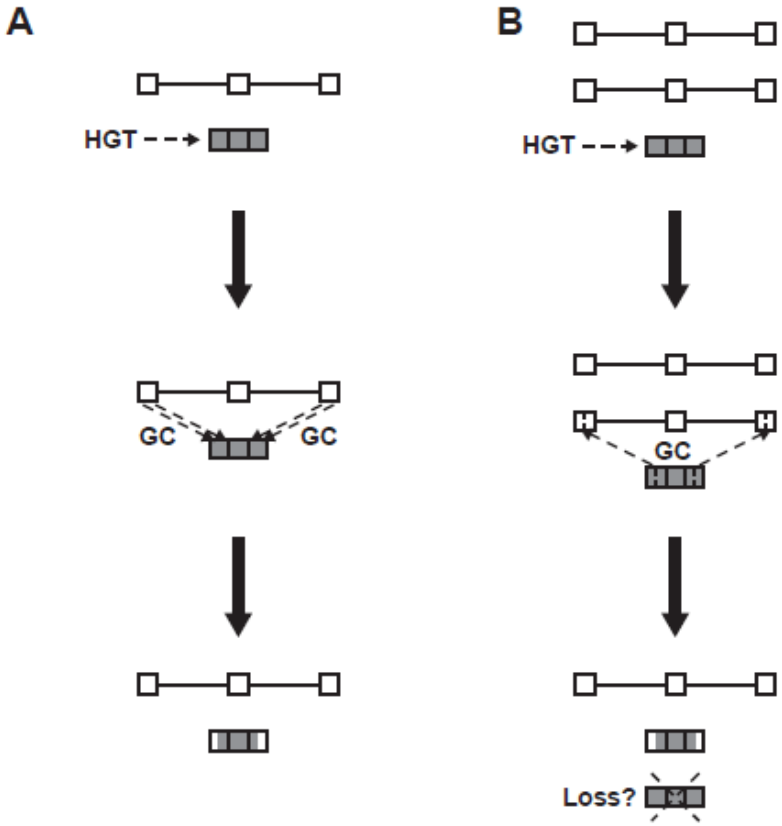


Figure 5. Models of intron loss via horizontal gene transfer (HGT) and gene conversion (GC). A) Model indicating HGT of an intronless *cox2* gene (in gray) followed by gene conversion of the extremities of the foreign gene by the native gene (in white). B) Model indicating HGT of an intronless *cox2* gene followed by gene conversion of one copy of the native intron-containing gene by the HGT gene. The HGT gene may be retained or subsequently lost.

ACKNOWLEDGEMENTS

We thank Richard Figlar and Laurence Ballard for providing leaf tissue for many of the *Magnolia* species used in this study; Sasha Kapil for helping with PCR surveys to assess *Magnolia* intron status; Yizhong Zhang for independently extracting *M. tripetala* DNA; Aaron Richardson, Greg Young, and Jeff Palmer for providing *Liriodendron* and *Amborella cox2* sequences prior to publication; Heather Achterberg for preparing voucher specimens; Felix Grewe and Alan Christensen for critical reading of the manuscript; and members of the Sally Mackenzie lab and the Mower lab for helpful discussions. This work was supported by start-up funds (JPM) and a UCARE award (DWS) from the University of Nebraska-Lincoln and by National Science Foundation awards IOS-1027529 and MCB-1125386 (JPM).

RESEARCH CONTRIBUTIONS

N.J. Hepburn generated most of the *cox2* data, performed the research, contributed with some figures and wrote the Material and Methods, all under the supervision of J.P. Mower. D.W. Schmidt performed all of the RNA extractions, amplified all of the cDNA sequences that were used in the RNA phylogenetic analysis, and carried out two RT-PCR assays under the supervision of N.J. Hepburn and J.P Mower. J.P. Mower conceived the study, performed the gene conversion analyses, prepared the figures and drafted the manuscript. All authors have read and approved the final version of the manuscript.

SUPPLEMENTARY TABLES

Supplementary Table 1. Source of material for plants used in this study.

Group	Order	Family	Species	Source [1]	Voucher
Magnoliids	Magnoliales	Annonaceae	*Asimina triloba*	EGMA	Achterberg & Mower 1002
	Magnoliales	Magnoliaceae	*Magnolia acuminata*	EGMA	Achterberg & Mower 1007
	Magnoliales	Magnoliaceae	*Magnolia cavaleriei* var. platypetala	JCRA	Achterberg & Mower 2012
	Magnoliales	Magnoliaceae	*Magnolia changhungtana*	JCRA	Achterberg & Mower 2020
	Magnoliales	Magnoliaceae	*Magnolia coco*	JCRA	Achterberg & Mower 2002
	Magnoliales	Magnoliaceae	*Magnolia conifera* var. chingii	JCRA	Achterberg & Mower 2015
	Magnoliales	Magnoliaceae	*Magnolia delavayi*	JCRA	Achterberg & Mower 2007
	Magnoliales	Magnoliaceae	*Magnolia ernestii*	JCRA	Achterberg & Mower 2010
	Magnoliales	Magnoliaceae	*Magnolia fordiana*	JCRA	Achterberg & Mower 2004
	Magnoliales	Magnoliaceae	*Magnolia insignis*	JCRA	Achterberg & Mower 2011
	Magnoliales	Magnoliaceae	*Magnolia kobus*	EGMA	Achterberg & Mower 1009
	Magnoliales	Magnoliaceae	*Magnolia kwangtungensis*	JCRA	Achterberg & Mower 2003
	Magnoliales	Magnoliaceae	*Magnolia laevifolia*	JCRA	Achterberg & Mower 2005
	Magnoliales	Magnoliaceae	*Magnolia liliifera*	JCRA	Achterberg & Mower 2001
	Magnoliales	Magnoliaceae	*Magnolia lotungensis*	JCRA	Achterberg & Mower 2018
	Magnoliales	Magnoliaceae	*Magnolia maudiae*	JCRA	Achterberg & Mower 2013
	Magnoliales	Magnoliaceae	*Magnolia odora*	JCRA	Achterberg & Mower 2006
	Magnoliales	Magnoliaceae	*Magnolia pyramidata*	NNN	Catalog # 113
	Magnoliales	Magnoliaceae	*Magnolia sinica*	JCRA	Achterberg & Mower 2016
	Magnoliales	Magnoliaceae	*Magnolia stellata*	EGMA	Achterberg & Mower 1008
	Magnoliales	Magnoliaceae	*Magnolia tamaulipana*	JCRA	Achterberg & Mower 2014
	Magnoliales	Magnoliaceae	*Magnolia tripetala*	NNN	Catalog # 115
	Magnoliales	Magnoliaceae	*Magnolia tripetala*	EGMA	Achterberg & Mower 1071
	Magnoliales	Magnoliaceae	*Magnolia virginiana* var. australis	JCRA	Achterberg & Mower 2017
	Magnoliales	Magnoliaceae	*Magnolia yuyuanensis*	JCRA	Achterberg & Mower 2009
	Laurales	Calycanthaceae	*Calycanthus floridus*	EGMA	Achterberg & Mower 1026
	Laurales	Lauraceae	*Persea americana*	BC	Achterberg & Mower 1112
	Piperales	Aristolochiaceae	*Aristolochia elegans*	BC	Achterberg & Mower 1131
	Piperales	Piperaceae	*Peperomia caperata*	BC	none
	Piperales	Saururaceae	*Houttuynia cordata*	BC	Achterberg & Mower 1145
Core Eudicots	Geraniales	Geraniaceae	*Geranium sanguineum*	BC	Achterberg & Mower 1152
Basal Eudicots	Buxales	Buxaceae	*Buxus microphylla*	EGMA	Achterberg & Mower 1025
	Proteales	Platanaceae	*Platanus* x *acerifolia*	EGMA	Achterberg & Mower 1057
	Ranunculales	Berberidaceae	*Mahonia repens*	EGMA	Achterberg & Mower 1067
	Ranunculales	Ranunculaceae	*Clematis integrifolia*	EGMA	Achterberg & Mower 1010
Basal Angiosperms	Austrobaileyales	Schisandraceae	*Schisandra chinensis*	BC	none

[1] Beadle Center, University of Nebraska-Lincoln (BC); Earl G. Maxwell Arboretum, University of Nebraska-Lincoln (EGMA); JC Raulston Arboretum, NC State University (JCRA); Nearly Native Nursery, Fayetteville, GA (NNN).

Supplementary Table 1. Source of plant materials

Supplementary Table 2. Primers used in this study

Primer name	Primer sequence (5`-3`)
PCR and RT-PCR primers	
cox2-e1F	TTGTGATGCWGCGGAACC
cox2-e2F	CGTTTATTAGAAGTNGACAATMGAGT
cox2-e2R	ACCTRAGGAAGGTACAGCC
cox2-e3R	GAGGATTAATTGATTGRATACCCR
cox2-i1F	CTACAACTTCRCCGAGCC
cox2-i2R1	CYACTAACCACTCGCTCG
cox2-i2R2	TGCCGTTGCAGGTCCTTC
qPCR primers	
cox2-Qe1F	CCCGATGTTCATTGCTATACC
cox2-Qe2R	GGAACTGTTATAGTCCGAATACTC
cox2-Qe2F	GCTGTACCTGGTCGTTTAAATC
cox2-Qe3R	CTTTCAAAGAAACTGCTTCTACG
cox2-Qi1F	TGACAGGATAGGTGAGGAATC
cox2-Qi1R	CAGTGGAAAGGGATGCTAG
cox2-Qi2F	CGAAGAGCTTAGTGTGAAACG
cox2-Qi2R	GCTACTCTTCCGTTTATCTCAC
cp-matK-QF	ATGGAAATCTCACCTTGTCC
cp-matK-QR	CGTACACTTGAAAGATAGCCC
cp-ndhF-QF	GGTGGGAATGTGTTCGTATC
cp-ndhF-QR	AACCAAAATCCCCTACACG
mt-atp1-QF	GGGAGTACCTATTGATGGAAG
mt-atp1-QR	CCTGTTTGCATAGGTTCGTG
mt-rpl10-QF	TGTTACGAGTAAGCGGAGAG
mt-rpl10-QR	GCAACTTGGTCGAAATGGG
nc-Gai1-QF	GCAGGTAGGGTGGAAATTAG
nc-Gai1-QR	TGAATTAACCGCCACAGC
nc-Lfy-QF	AGAGCTTTCAAGGAGAGGG
nc-Lfy-QR	ACATACCATATAGTGAGACGGG
nc-PhyA-QF	GCAGTCGAAGAGGAAGAGG
nc-PhyA-QR	CGTAGAGGGAAAGGAACAAAC

Supplementary Table 2. Primer sequences

Supplementary Table 3. GenBank accession numbers for plants used in this study

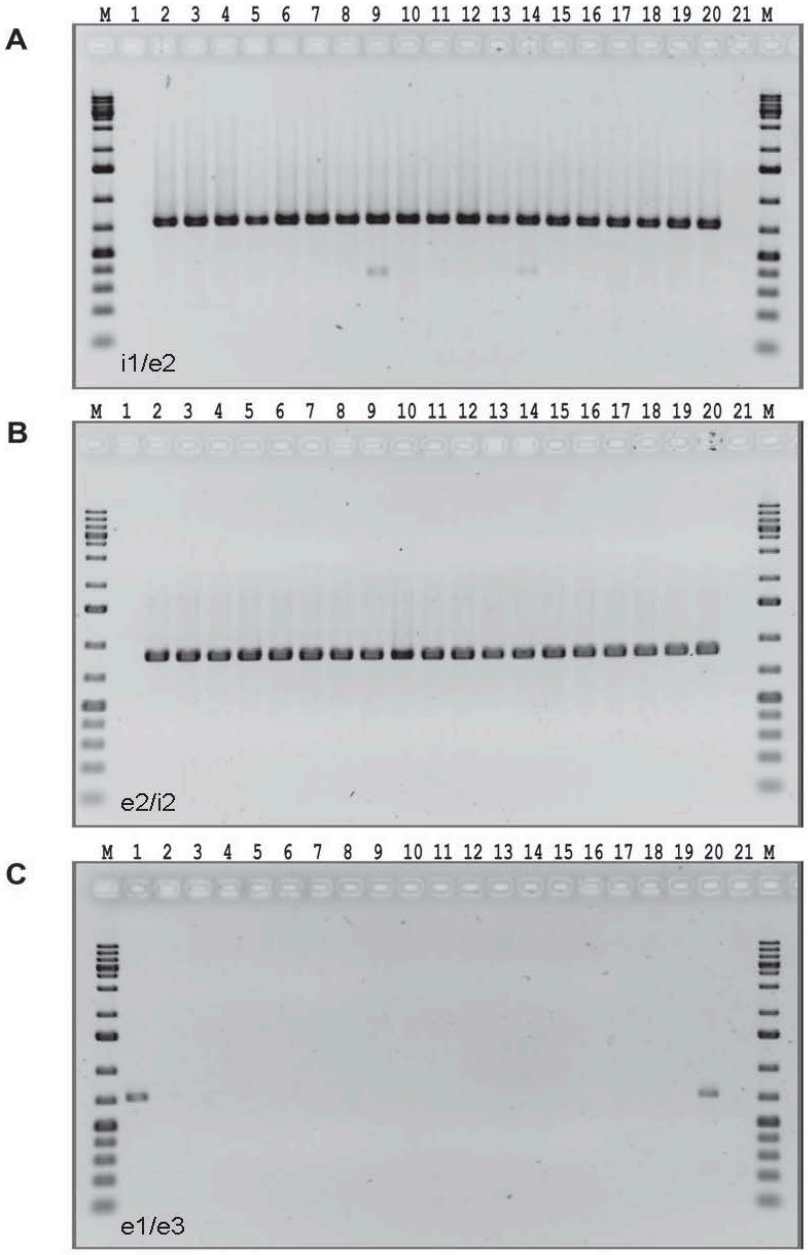
Group	Order	Family	Species	Intron Status	Genbank Accession
Magnoliids	Magnoliales	Annonaceae	*Asimina triloba*	(+/+)	JQ317142
	Magnoliales	Magnoliaceae	*Liriodendron tulipifera*	(+/+)	unpublished data [1]
	Magnoliales	Magnoliaceae	*Magnolia acuminata*	(+/+)	JQ317147
	Magnoliales	Magnoliaceae	*Magnolia cavaleriei var. platypetala*	(+/+)	JQ317148
	Magnoliales	Magnoliaceae	*Magnolia fordiana*	(+/+)	JQ317146
	Magnoliales	Magnoliaceae	*Magnolia kobus*	(+/+)	JQ317149
	Magnoliales	Magnoliaceae	*Magnolia liliifera*	(+/+)	JQ317145
	Magnoliales	Magnoliaceae	*Magnolia pyramidata* (GenBank)	(−/−)	U42696
	Magnoliales	Magnoliaceae	*Magnolia pyramidata* (NNN)	(+/+)	JQ317144
	Magnoliales	Magnoliaceae	*Magnolia stellata*	(+/+)	JQ317150
	Magnoliales	Magnoliaceae	*Magnolia tamaulipana*	(+/+)	JQ317151
	Magnoliales	Magnoliaceae	*Magnolia tripetala* (EGMA)	(−/−)	JQ317133
	Magnoliales	Magnoliaceae	*Magnolia tripetala* (EGMA)	(+/+)	JQ317152
	Magnoliales	Magnoliaceae	*Magnolia tripetala* (Genbank)	(−/−)	U43055
	Magnoliales	Magnoliaceae	*Magnolia tripetala* (NNN)	(−/−)	JQ317132
	Laurales	Calycanthaceae	*Calycanthus floridus*	(+/+)	JQ317141
	Laurales	Lauraceae	*Laurus nobilis*	(+/?)	AY832094, AY832104
	Laurales	Lauraceae	*Persea americana*	(+/+)	JQ317143
	Piperales	Aristolochiaceae	*Aristolochia elegans*	(+/+)	JQ317137
	Piperales	Aristolochiaceae	*Asarum sp.*	(+/?)	AY832096, AY832105
	Piperales	Piperaceae	*Peperomia caperata*	(+/+)	JQ317135
	Piperales	Piperaceae	*Piper betle*	(+/+)	AY832091, AY832102
	Piperales	Saururaceae	*Houttuynia cordata*	(+/+)	JQ317139
Monocots	Asparagales	Amaryllidaceae	*Allium cepa*	(+/−)	GU253305,GU253307
	Asparagales	Asparagaceae	*Agave attenuata*	(+/?)	AY832092, AY832103
	Commelinales	Pontederiaceae	*Eichhornia crassipes*	(+/?)	AY832093, AY832099
	Poales	Poaceae	*Oryza sativa*	(+/−)	DQ167399
	Poales	Poaceae	*Zea mays*	(+/−)	AY506529
Core Eudicots	Apiales	Apiaceae	*Daucus carota*	(+/+)	X63625
	Brassicales	Brassicaceae	*Arabidopsis thaliana*	(−/+)	Y08501
	Brassicales	Caricaceae	*Carica papaya*	(−/+)	EU431224
	Caryophyllales	Amaranthaceae	*Beta vulgaris*	(+/−)	BA000009
	Caryophyllales	Caryophyllaceae	*Silene latifolia*	(−/−)	NC_014487
	Cucurbitales	Cucurbitaceae	*Citrullus lanatus*	(−/+)	GQ856147
	Cucurbitales	Cucurbitaceae	*Cucurbita pepo*	(−/+)	GQ856148
	Fabales	Fabaceae	*Pisum sativum*	(−/−)	AJ414385
	Geraniales	Geraniaceae	*Geranium sanguineum*	(−/−)	JQ317153
	Lamiales	Phrymaceae	*Mimulus guttatus*	(−/+)	JN098455
	Malpighiales	Euphorbiaceae	*Ricinus communis*	(−/+)	HQ874649
	Myrtales	Onagraceae	*Oenothera villaricae*	(−/−)	X00212
	Solanales	Solanaceae	*Capsicum annuum*	(+/−)	DQ126683
	Solanales	Solanaceae	*Nicotiana tabacum*	(+/−)	BA000042
	Vitales	Vitaceae	*Vitis vinifera*	(+/+)	FM179380
Basal Eudicots	Buxales	Buxaceae	*Buxus microphylla*	(+/+)	JQ317138
	Proteales	Platanaceae	*Platanus x acerifolia*	(+/+)	JQ317140
	Ranunculales	Berberidaceae	*Mahonia repens*	(−/−)	JQ317131
	Ranunculales	Ranunculaceae	*Clematis integrifolia*	(−/+)	JQ317134
Basal Angiosperms	Amborellales	Amborellaceae	*Amborella trichopoda*	(+/+)	unpublished data [1]
	Austrobaileyales	Schisandraceae	*Schisandra chinensis*	(+/+)	JQ317136

[1] unpublished data provided by AO Richardson, GJ Young, and JD Palmer

Supplementary Table 3. GenBank accession numbers

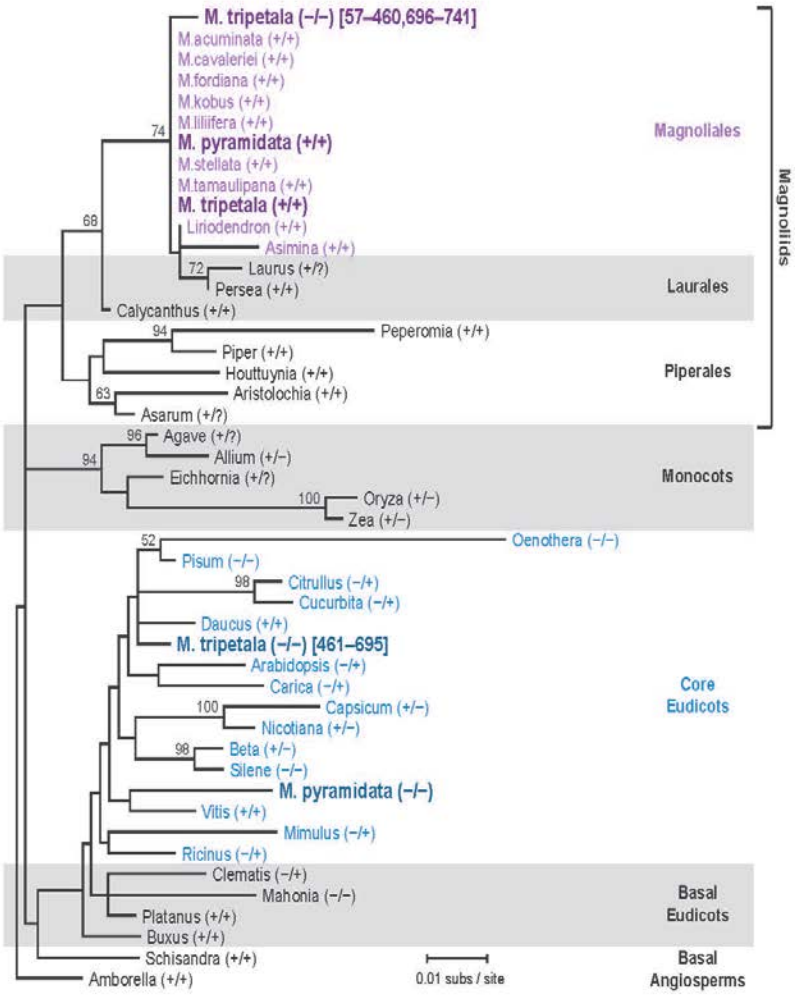
SUPPLEMENTARY FIGURES

Supplementary Figure 1. Intron status of the *cox2* gene in additional *Magnolia* species. Assay results using primers shown in Figure 1 designed to amplify a copy of *cox2* that A) contains intron 1, B) contains intron 2, or C) lacks introns. Lane 1: *Geranium sanguinium*, 2: *Magnolia lilifera*, 3: *M. coco*, 4: *M. kwangtungensis*, 5: *M. fordiana*, 6: *M. laevifolia*, 7: *M. odora*, 8: *M. delavayi*, 9: *M. yuyuanensis*, 10: *M. ernestii*, 11: *M. insignis*, 12: *M. cavaleriei*, 13: *M. maudiae*, 14: *M. tamaulipana*, 15: *M. confifera*, 16: *M. sinica*, 17: *M. virginiana*, 18: *M. lotungensis*, 19: *M. changhungtana*, 20: *M. tripetala*, 21: negative control, M: GeneRuler 1 kb plus DNA ladder (Fermentas).



Supplementary Figure 1. Intron status of the *cox2* gene in additional *Magnolia* species.

Supplementary Figure 2. Phylogenetic analysis of the Magnoliales-like portion (positions 57–460 and 696–741) and the eudicot-like portion (positions 461–695) of the *M. tripetala* intronless *cox2* gene. The *cox2* intron status is given to the right of each species name. RNA sequences were used for all species. Bootstrap values from 1000 replicates are shown when >50%. Tree was rooted on basal angiosperms. M. = *Magnolia*.



Supplementary Figure 2. Phylogenetic analysis of *cox2* RNA sequences

REFERENCES

Albrizio, M, L De Gara, C De Benedetto, O Arrigoni, R Gallerani. 1994. Investigations of the *coxII* intron structure in the mitochondrial genes of angiosperms. Plant Sci 100:179-186.

Barkman, TJ, JR McNeal, SH Lim, G Coat, HB Croom, ND Young, CW Depamphilis. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. BMC Evol Biol 7:248.

Belshaw, R, D Bensasson. 2006. The rise and falls of introns. Heredity 96:208-213.

Bergthorsson, U, KL Adams, B Thomason, JD Palmer. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 424:197-201.

Bock, R. 2010. The give-and-take of DNA: horizontal gene transfer in plants. Trends Plant Sci 15:11-22.

Bonen, L. 2011. RNA Splicing in Plant Mitochondria. In: F Kempken, editor. Plant Mitochondria: Springer New York. p. 131-155.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540-552.

Cho, Y, Y-L Qiu, P Kuhlman, JD Palmer. 1998. Explosive invasion of plant mitochondria by a group I intron. Proc Natl Acad Sci U S A 95:14244-14249.

Coulombe-Huntington, J, J Majewski. 2007. Intron loss and gain in *Drosophila*. Mol Biol Evol 24:2842-2850.

De Benedetto, C, L De Gara, O Arrigoni, M Albrizio, R Gallerani. 1992. The structure of the cytochrome oxidase subunit II gene and its use as a new character in the construction of the phylogenetic tree of Angiospermae. Plant Sci 81:75-82.

Derr, LK, JN Strathern. 1993. A role for reverse transcripts in gene conversion. Nature 361:170-173.

Dombrovska, O, Y-L Qiu. 2004. Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. Mol Phylogenet Evol 32:246-263.

Draper, CK, JB Hays. 2000. Replication of chloroplast, mitochondrial and nuclear DNA during growth of unirradiated and UVB-irradiated *Arabidopsis* leaves. Plant J 23:255-265.

Edgar, RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797.

Fink, GR. 1987. Pseudogenes in yeast? Cell 49:5-6.

Geiss, KT, GM Abbas, CA Makaroff. 1994. Intron loss from the NADH dehydrogenase subunit 4 gene of lettuce mitochondrial DNA: evidence for homologous recombination of a cDNA intermediate. Mol Gen Genet 243:97-105.

Grewe, F, S Herres, P Viehöver, M Polsakiewicz, B Weisshaar, V Knoop. 2011. A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. Nucleic Acids Res 39:2890-2902.

Guindon, S, JF Dufayard, V Lefort, M Anisimova, W Hordijk, O Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307-321.

Hall, TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95-98.

Hao, W. 2010. OrgConv: detection of gene conversion using consensus sequences and its application in plant mitochondrial and chloroplast homologs. BMC Bioinformatics 11:114.

Hao, W, JD Palmer. 2009. Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. Proc Natl Acad Sci U S A 106:16728-16733.

Hao, W, AO Richardson, Y Zheng, JD Palmer. 2010. Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. Proc Natl Acad Sci U S A 107:21576-21581.

Hecht, J, F Grewe, V Knoop. 2011. Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. Genome Biol Evol 3:344-358.

Hiesel, R, A Brennicke. 1983. Cytochrome oxidase subunit II gene in mitochondria of *Oenothera* has no intron. EMBO J 2:2173-2178.

Itchoda, N, S Nishizawa, H Nagano, T Kubo, T Mikami. 2002. The sugar beet mitochondrial nad4 gene: an intron loss and its phylogenetic implication in the Caryophyllales. Theor Appl Genet 104:209-213.

Joly, S, L Brouillet, A Bruneau. 2001. Phylogenetic implications of the multiple losses of the mitochondrial *coxII.i3* intron in the angiosperms. Int J Plant Sci 162:359-373.

Kubo, T, S Nishizawa, A Sugawara, N Itchoda, A Estiati, T Mikami. 2000. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for $tRNA^{Cys}$(GCA). Nucleic Acids Res 28:2571-2576.

Kudla, J, FJ Albertazzi, D Blazevic, M Hermann, R Bock. 2002. Loss of the mitochondrial *cox2* intron 1 in a family of monocotyledonous plants and utilization of mitochondrial intron sequences for the construction of a nuclear intron. Mol Genet Genomics 267:223-230.

Lambowitz, AM, M Belfort. 1993. Introns as mobile genetic elements. Annu Rev Biochem 62:587-622.

Lambowitz, AM, S Zimmerly. 2011. Group II introns: Mobile ribozymes that invade DNA. Cold Spring Harbor Perspectives in Biology 3.

Lamppa, GK, AJ Bendich. 1984. Changes in mitochondrial DNA levels during development of pea (*Pisum sativum* L.). Planta 162:463-468.

Lang, BF, MJ Laforest, G Burger. 2007. Mitochondrial introns: a critical view. Trends Genet 23:119-125.

Llopart, A, JM Comeron, FG Brunet, D Lachaise, M Long. 2002. Intron presence-absence polymorphism in *Drosophila* driven by positive Darwinian selection. Proc Natl Acad Sci U S A 99:8121-8126.

Mourier, T, DC Jeffares. 2003. Eukaryotic intron loss. Science 300:1393.

Mower, JP. 2009. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res 37:W253-259.

Mower, JP, AL Case, ER Floro, JH Willis. in press. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. Genome Biol Evol.

Mower, JP, DB Sloan, AJ Alverson. in press. Plant mitochondrial genome diversity: the genomics revolution. In: JH Wendel, editor. Plant Genome Diversity Volume 1: Plant Genomes, their Residents, and their Evolutionary Dynamics. New York: Springer.

Mower, JP, S Stefanović, W Hao, JS Gummow, K Jain, D Ahmed, JD Palmer. 2010. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. BMC Biol 8:150.

Parma, J, D Christophe, V Pohl, G Vassart. 1987. Structural organization of the 5' region of the thyroglobulin gene. Evidence for intron loss and "exonization" during evolution. J Mol Biol 196:769-779.

Pruitt, KD, MR Hanson. 1989. Cytochrome oxidase subunit II sequences in *Petunia* mitochondria: two intron-containing genes and an intron-less pseudogene associated with cytoplasmic male sterility. Curr Genet 16:281-291.

Qiu, Y-L, YR Cho, JC Cox, JD Palmer. 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394:671-674.

Rabbi, MF, KG Wilson. 1993. The mitochondrial *coxII* intron has been lost in two different lineages of dicots and altered in others. Am J Bot 80:1216-1223.

Ran, JH, H Gao, XQ Wang. 2010. Fast evolution of the retroprocessed mitochondrial *rps3* gene in Conifer II and further evidence for the phylogeny of gymnosperms. Mol Phylogenet Evol 54:136-149.

Richardson, AO, JD Palmer. 2007. Horizontal gene transfer in plants. J Exp Bot 58:1-9.

Robbens, S, E Derelle, C Ferraz, J Wuyts, H Moreau, Y Van de Peer. 2007. The complete chloroplast and mitochondrial DNA sequence of *Ostreococcus tauri*: organelle genomes of the smallest eukaryote are examples of compaction. Mol Biol Evol 24:956-968.

Roy, SW, W Gilbert. 2005. The pattern of intron loss. Proc Natl Acad Sci U S A 102:713-718.

Sanchez-Puerta, MV, Y Cho, JP Mower, AJ Alverson, JD Palmer. 2008. Frequent, phylogenetically local horizontal transfer of the *cox1* group I Intron in flowering plant mitochondria. Mol Biol Evol 25:1762-1777.

Schmitz-Linneweber, C, A Barkan. 2007. RNA splicing and RNA editing in chloroplasts. In: R Bock, editor. Cell and Molecular Biology of Plastids: Springer Berlin / Heidelberg. p. 213-248.

Sloan, DB, AJ Alverson, H Storchova, JD Palmer, DR Taylor. 2010a. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. BMC Evol Biol 10:274.

Sloan, DB, AH MacQueen, AJ Alverson, JD Palmer, DR Taylor. 2010b. Extensive loss of RNA editing sites in rapidly evolving *Silene* mitochondrial genomes: selection vs. retroprocessing as the driving force. Genetics 185:1369-1380.

Sugiyama, Y, Y Watase, M Nagase, N Makita, S Yagura, A Hirai, M Sugiura. 2005. The complete nucleotide sequence and multipartite organization of the tobacco

mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol Genet Genomics 272:603-615.

Wang, W, H Zheng, S Yang, et al. 2005. Origin and evolution of new exons in rodents. Genome Res 15:1258-1264.

CHAPTER 3

Mechanisms of group II intron loss from the mitochondrial *cox2* gene of plants

Nancy J. Hepburn[1,2], Derek W. Schmidt[1,3], and Jeffrey P. Mower[1,2]

[1] Center for Plant Science Innovation, [2] Department of Agronomy and Horticulture, and [3] School of Biological Sciences, University of Nebraska, Lincoln, NE 68588, USA

ABSTRACT

Eukaryotic evolution has been characterized by intron loss. In eukaryotes, the proposed mechanisms of intron loss are exonization, random genomic deletion, retroprocessing and most recently, horizontal gene transfer followed by gene conversion (HGT-GC). Here we investigate the mechanism of intron loss from the mitochondrial *cox2* gene, which in previous studies has shown a variable intron distribution in plants. Through an extensive PCR survey that included 107 tracheophyte species, gene sequencing, and phylogenetic analyses, we determined that the variable intron distribution is better explained by intron loss rather than intron gain. We found no evidence supporting intron loss via exonization and random genomic deletion, while limited data was consistent with retroprocessing.

Interestingly, the *cox2* exon phylogeny did not fully recover organismal relationships, an indication of the effects of horizontal gene transfer. Four magnoliid intronless paralogs showed strong phylogenetic conflicts, and several members of other clades (*Acorus, Ruscus*, rosids, and asterids) were also found in unexpected positions, although with weaker support. Further research is needed to determine if these intronless paralogs arose via HGT-GC and to determine the role that HGT-GC has played in intron loss dynamics.

INTRODUCTION

The organellar genomes of land plants are often interrupted by group I and II introns. Group II introns are particularly abundant in the mitochondrial genomes of angiosperms (Bonen 2008). In angiosperms, the evolutionary history of most mitochondrial group II introns is quite stable. However, the group II introns located in the *cox2* gene are unusual in this regard since they have been repeatedly and independently lost across different lineages over time (Joly, Brouillet, Bruneau 2001; Kudla et al. 2002; Hepburn, Schmidt, Mower 2012). Thus, they serve as a great model for understanding the dynamics of intron loss.

In plants, there are two described models for the biological mechanism of intron loss: random chromosomal deletion and retroprocessing (reviewed in detail in Chapters 1 and 2). Nevertheless, in 2012 we proposed a novel mechanism of intron loss, which involved horizontal gene transfer (HGT) and gene conversion (GC) (Hepburn, Schmidt, Mower 2012). This mechanism involves the horizontal transfer of an intronless gene that undergoes GC with the native intron containing copy, giving rise to a chimeric intronless gene. The latest model is plausible because HGT is a well documented and fairly common phenomenon in plant mitochondria (Nickrent et al. 1998; Bergthorsson et al. 2003; Won, Renner 2003; Bergthorsson et al. 2004; Davis, Wurdack 2004; Mower et al. 2004; Woloszynska et al. 2004; Davis, Anderson, Wurdack 2005; Richardson, Palmer 2007; Keeling, Palmer 2008; Hao et al. 2010; Mower et al. 2010). Furthermore, reports of horizontally acquired genes undergoing gene conversion with the native genes have emerged (Bergthorsson et al. 2003; Barkman et al. 2007; Richardson, Palmer 2007; Hao, Palmer 2009; Hao et al. 2010; Mower et al. 2010; Hepburn, Schmidt, Mower 2012). Perhaps the delay in reported cases is due, in part, because the combined effects of

HGT and GC can be difficult to detect, especially if the transferred or converted segment is small.

In this chapter, I describe our efforts to identify which of the proposed mechanisms of intron loss is responsible for the widespread loss of introns from the *cox2* gene. Through statistical and phylogenic work, it becomes apparent that there are several mechanisms responsible for the observed intron distribution in the *cox2* gene of plants. We find some clear cases of retroprocessing, but we also find more evidence to support our own proposed HGT-GC mechanism. In contrast, we find no evidence for genomic deletion or for exonization.

Materials and Methods

Plant Materials

Fresh leaves from a total of 107 samples representing three major vascular plant groups (monilophytes, gymnosperms and angiosperms) were obtained from the greenhouse at the Beadle Center and from the Earl G. Maxwell Arboretum at the University of Nebraska-Lincoln. Voucher information for most selected plants is provided in Table S1.

Nucleic acid extraction, gene amplification & sequencing

Total DNA and RNA samples were isolated from fresh leaf tissue as described by Hepburn et al. (2012). Several primer pair combinations were used to amplify the *cox2* gene for these diverse plants. Most angiosperm species were amplified using degenerate *cox2* primers labeled as UNI (Table S2). When these primer combinations failed, “EMBRYO” primers were used to amplify the gene. Gymnosperms and monilophytes were amplified using a combination of specific primers (GYMNO or PTERI respectively) in conjunction with degenerate primers (Table S2). To completely

sequence both introns for all species, primer walking was performed using primers from Table S3 and Table S4.

All PCR reactions were initially carried out in a 60 ul reaction mixture using primer pairs from Table S2, GoTaq Flexi DNA Polymerase, and supplied reagents (Promega, WI, USA) following the manufacturer's instructions. Each reaction was amplified on an Engine Dyad Peltier Thermal Cycler-0220 or in a C1000 Thermal Cycler (Bio-Rad, CA, USA). The PCR program parameters for all described assays included a pre-denaturation step of 94ºC for 3 min, followed by 35 cycles of denaturation at 94ºC for 30 s, annealing at 50ºC for 45 s, and elongation at 72ºC for 2 min, with a final step of 72ºC for 5 min.

All amplicons were purified and sequenced on both strands at the High-Throughput Genomics Unit (University of Washington, Seattle, USA). All newly generated sequences will be deposited in GenBank.

cDNA synthesis, PCR amplification & RNA editing analysis

cDNAs were synthesized and amplified as previously described (Hepburn, Schmidt, Mower 2012). Sites of RNA editing were determined experimentally for 46 species by comparing the DNA and cDNA sequences. For the rest of the species, edit sites were predicted using PREP-Aln (Mower 2009) with the detection cutoff value set to 0.2.

Taxon sampling, alignments & phylogenetic analyses

The mitochondrial *cox2* gene was amplified from 107 species, while an additional 61 *cox2* DNA sequences were retrieved from GenBank. The compiled dataset included 168 species representing 10 major groups within the tracheophytes (Table S1).

Sequences were assembled in CodonCode Aligner version 3.5 (CodonCode Corp., Dedham, MA) and sequence alignments were performed using MUSCLE inside the program MEGA version 5.0 (Edgar 2004; Tamura et al. 2011). Sampling was aimed to be as comprehensive as possible, although partial coding sequences (< 500 bp) were excluded from the alignments. Sequence alignments were manually refined using BioEdit version 5.0.6 (Hall 1999). Poor-quality alignment regions were removed by GBlocks version 0.91b (Talavera, Castresana 2007) and the parameters were set to be less stringent (by selecting the three options of the online version).

All phylogenetic analyses were performed with phyML online version 3.0 (Guindon et al. 2010) using the Maximum Likelihood (ML) algorithm. The *cox2* gene tree was built exclusively from cDNA sequences. In all the phylogenetic analyses the trees were rooted on Lycophytes. The general time-reversible (GTR) model was used as the substitution model. Base frequencies were set to empirical, while the proportion of invariable sites and the shape of the gamma distribution with four substitution rate categories (GTR+G+I) were estimated during the run. Tree improvement was performed by subtree pruning and regrafting with ten random starting trees. Tree support was evaluated by bootstrapping from 1000 replicates.

Testing for character correlations & stochastic character mapping

To determine if there were any character correlations between the number of introns and edit sites, we performed stochastic mapping simulations using SIMMAP Version 1.5 (Bollback 2006). SIMMAP is a Bayesian approach based on stochastic models that uses mutational mapping that is consistent with the distribution across the tips of the topology to estimate the posterior probability distribution from several ancestral states reconstructions, also known as stochastic realizations (Ronquist 2004; Bollback 2006).

Correlation analyses were performed using a cDNA maximum likelihood tree built from a reduced data set containing 107 species. Species were excluded from the analysis if they were short (since the predicted number of edit sites would be misleading), if the presence or absence of both introns was not known, and if there was an overrepresentation of closely related individuals that contained the same intron number and similar number of edit sites. The total number of edit sites was coded into three main categories: 0, 1 and 2. Category 0 included species possessing 0-6 edit sites, category 1 included species with 7-11 edit sites, and category 2 included species with 12 or more edit sites. Intron number was not modified. The stochastic mappings and correlation analyses were configured as a morphology/standard model. Intron number and edit site number characters were set to the empirical prior distribution, while the γ prior distribution was set to 90. Tree space was searched using Markov chain Monte Carlo algorithm (MCMC) with standard parameters.

SIMMAP version 1.5 determines the time spent on a particular character state. It compares the reconstructed states for two characters to determine if they covary more than expected by chance. For this particular analysis we tested for character correlation using D statistics, which is the overall difference between the observed and expected, based on simulations (Huelsenbeck, Nielsen, Bollback 2003; Bollback 2006). To check for the robustness of the results, the analysis was carried out 3 independent times using different number of prior draws and predictive sampling (45, 47, and 50).

RESULTS

Patchy cox2 *intron distribution in angiosperms*

Through an extensive *cox2* PCR survey (and gene sequencing), it became clear that there was significant variation in the intron distribution across angiosperms (Fig S0). Moreover, there was not an apparent correlation between the two introns, which suggested that the observed intron distribution was due to one of two things: several intron gains (as seen in the plant mitochondrial *cox1* gene) or frequent and independent intron losses.

The observed patchy distribution is due to intron loss rather than gain

Several studies have demonstrated that the group I intron located in the mitochondrial *cox1* gene of many angiosperms has been horizontally transferred across hundreds of species (Vaughn et al. 1995; Cho et al. 1998; Sanchez-Puerta et al. 2008; Sanchez-Puerta et al. 2011). It is thought that the intron has an intact and active homing endonuclease, which facilitates its spread (Sanchez-Puerta et al. 2011). Taking this into consideration, and the fact that some group I and II introns are mobile self-splicing ribozymes that can successfully spread (Lambowitz, Zimmerly 2011), we tested whether the observed intron distribution (Figure S0) was better explained by intron gain via HGT rather than intron loss.

If these introns have been horizontally acquired, the intron phylogeny will be highly incongruent, reflecting a pattern of horizontal transfer rather than organismal relationships. On the other hand, if these introns have been vertically inherited and lost stochastically, then the remaining introns should group consistently with known organismal relationships (Mower, Jain, Hepburn, in press).

Our maximum likelihood (ML) analysis for the intron 1 (i373) phylogeny was based on 83 taxa comprising 1048 characters (Figure 1). The result from this molecular phylogenetic analysis was overall in congruence with currently accepted organismal relationships. We were able to recover many expected monophyletic groups with strong to weak bootstrap support (Figure 1; bootstrap values shown in Figure S1). The bootstrap values for the monophyletic groups were as following: lycophytes (100%), gymnosperms (100%), caryophyllids (88%), asterids (69%), monocots (61%), magnoliids (53%) and rosids (35%). In addition, the two basal angiosperms *Amborella trichopoda* and *Schisandra chinensis* were placed appropriately as early diverging angiosperms. Similarly, the two basal eudicots (*Platanus x acerifolia* and *Buxus microphylla*) split as expected near the root of core eudicots. In contrast to the above congruent results, a magnoliid species (*Magnolia grandiflora)* was nested within the asterids, which strongly suggested HGT. However, this sequence was obtained from GenBank, so we cannot exclude the possibility of sample misidentification. In addition, Monilophytes were recovered as paraphyletic, although this result was not strongly supported by bootstrapping (47%).

Figure 1. Maximum likelihood phylogenetic tree resulting from the analysis of 83 *cox2* intron 1 sequences. Bootstrap values from 1000 replicates are shown in supplemental Figure S1. Species are color coded as following: monilophytes (orange), gymnosperms (blue), magnoliids (dark red), basal angiosperms (gold), monocots (magenta), basal eudicots (dim gray), rosids (navy), caryophyllids (turquoise blue), and asterids (forest green).

Figure 2 shows the ML phylogeny for *cox2* intron 2 (i691). The tree was built from a sequence alignment consisting of 78 sequences comprising 1264 characters. Just as in the previous phylogenetic analysis, the tree was generally consistent with known organismal relationships. We recovered several monophyletic groups with strong to weak bootstrap support (Figure S2), including lycophytes (100%), gymnosperms (100%), caryophyllids (100%), monocots (98%), and rosids (9%). Once again, basal angiosperms and basal eudicots appeared in their overall expected positions. Most magnoliids also appeared in the phylogeny as expected, but they were recovered in four

weakly supported paraphyletic groups. Asterids were found to be monophyletic except for *Chionanthus virginicus,* which was nested within the rosids, albeit with low bootstrap support. In two preliminary analyses, however, *C.virginicus* was recovered within the asterids with 100% bootstrap support, as the sister taxon to other Lamiales *(Digitalis purpurea and Mimulus gattutus*), suggesting that this intron was vertically acquired. It is likely that a misalignment occurred in this sequence during the final phylogenetic analysis, which resulted in a lower phylogenetic signal. Although caryophyllids were recovered as a monophyletic group (100% bootstrap support), they appeared nested within rosids with weak bootstrap support.

The general congruence of the intron phylogenies with known organismal relationships argues strongly against a model of numerous horizontal gains. Instead, the patchy intron distributions are most likely caused by frequent intron losses.

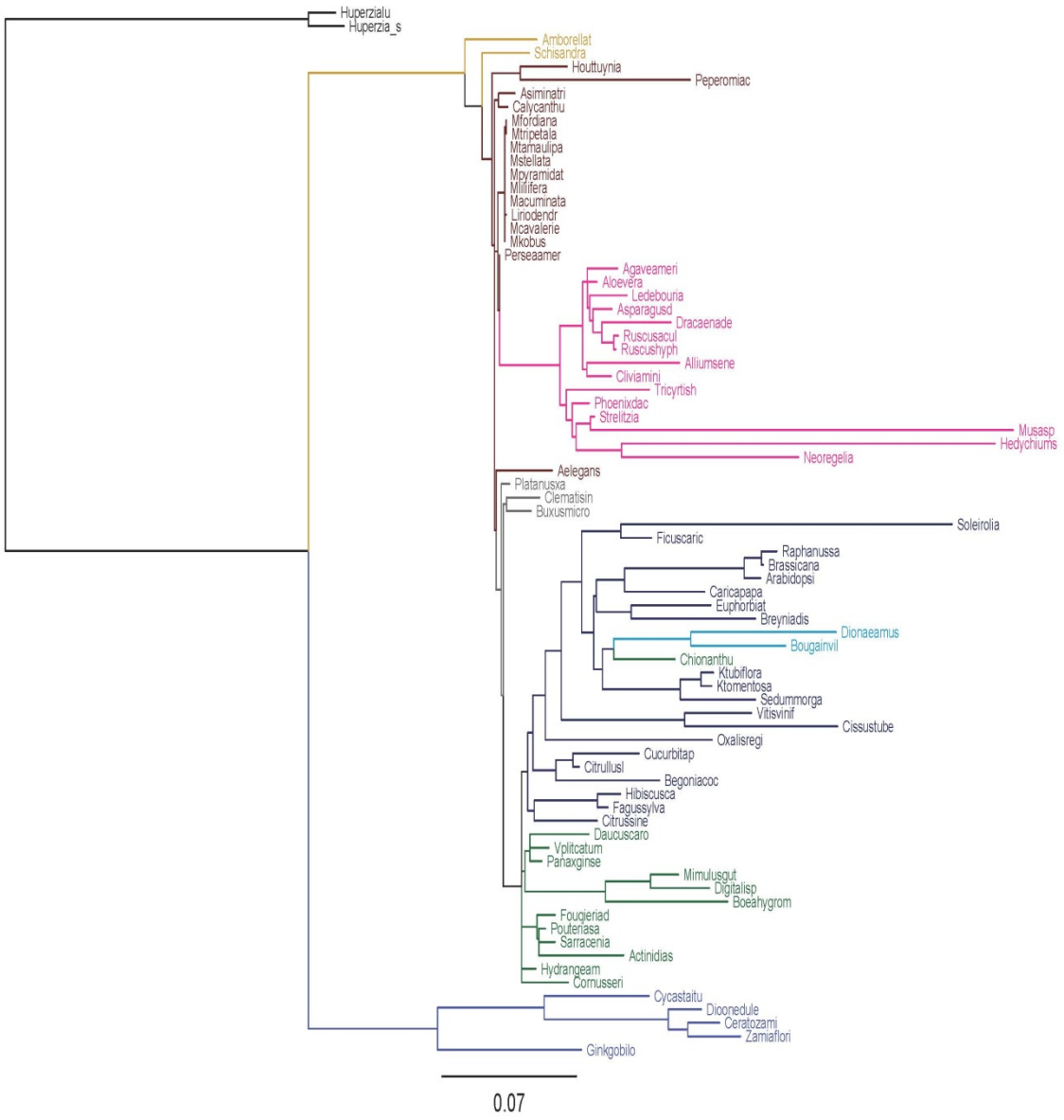


Figure 2. Maximum likelihood phylogenetic tree resulting from the analysis of 78 *cox2* intron 2 sequences. Bootstrap values from 1000 replicates are shown in supplemental Figure S2. Major groups were color coded as described in Fig.1.

The cox2 exon phylogeny is not congruent with current plant taxonomy

Figure 3 shows a ML phylogeny based on *cox2* cDNA sequences that included exon 1, 2 and 3. The data set included 687 characters from 141 taxa, representing 10 major groups within the tracheophytes (Table S1). The analysis successfully recovered four expected monophyletic groups with the following bootstrap support values: lycophytes (60%), monilophytes (39%), gymnosperms (99%), basal and caryophyllids (54%) (Figure S3). Interestingly, we found several phylogenetic conflicts, and a few of them were strongly supported (Figure 3 and S3). For instance, although most of the magnoliid species formed a monophyletic group (38% bootstrap support), four putative intronless paralogs were found in unexpected positions. *Magnolia cavaleriei's* paralog (McavXI) grouped with *Vinca minor* (an asterid) with strong support (100% bootstrap support). Similarly, *Magnolia kobus*' paralog (Mkob9pB46) appeared nested within a small asterid clade with 82% bootstrap support, while the paralog sequence from the magnoliid *Persea americana* (Pamerixii showed a strong phylogenetic affinity towards the sequence from the asterid *Pouteria sapota* (91% bootstrap support). Monocots were also split into 3 paraphyletic groups. The most species-rich group was recovered with 75% bootstrap support, whereas the smaller groups containing species from the orders Acorales and Asparagales did not have a strong affiliation to any particular clade. Rosids and asterids were scattered around the phylogeny; however, this result received little to no bootstrap support.

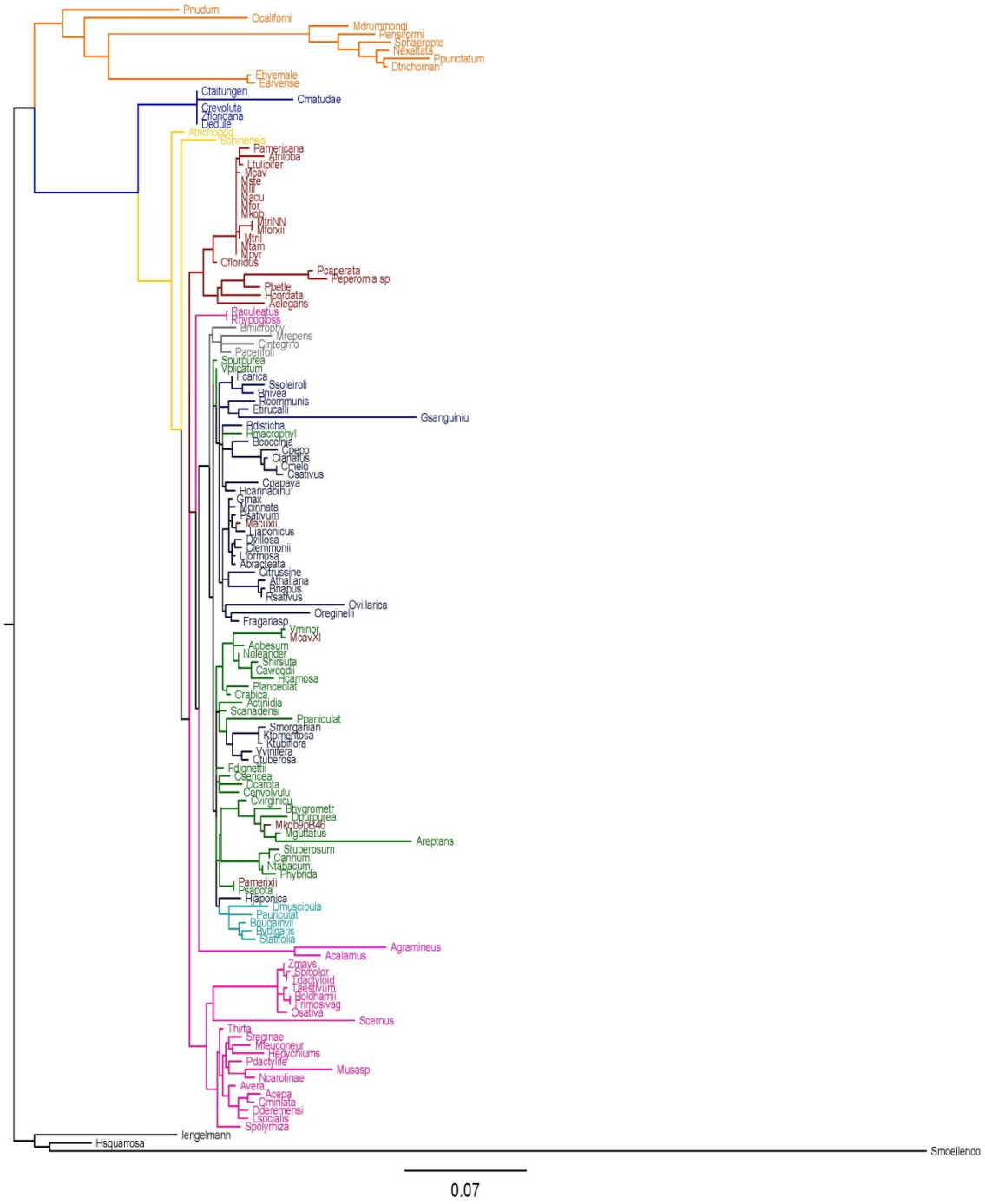


Figure 3. Maximum Likelihood phylogenetic tree resulting from the analysis of 141 *cox2* cDNA sequences. Bootstrap values from 1000 replicates are shown in supplemental Figure S3. Major groups are color coded as described in Fig. 1.

A weak correlation exists between intron loss and edit site loss

Three independent analyses were carried out using SIMMAP 1.5. The analyses differed by the predicted number of samples that were taken during the simulations. Table 1 shows the results of the correlation analyses between intron numbers and edit site numbers based on a reduced ML *cox2* cDNA phylogeny (Figure 4, bootstrap values are shown in Figure S4). Overall, there was a weak but positive correlation between intron number and edit site number (higher positive correlations are shown in green). I would like to point out, however, that the correlation between the absence of introns (coded as 0) and a low number of edit sites (coded as 0) was not significant. The only significant positive correlation was observed for the largest number of introns with the highest amount of edit sites (Table 1C).

Character 2 (edit site number)

Character 1 (intron number)

A

	0	1	2
0	**0.0206**	0.0169	-0.0234
1	0.0006	**0.0224**	-0.0164
2	-0.0104	-0.0231	**0.0542**

B

	0	1	2
0	**0.0221**	0.0160	-0.0238
1	0.0001	**0.0230**	-0.0166
2	-0.0110	-0.0230	**0.0549**

C

	0	1	2
0	**0.0163**	0.0124	-0.0188
1	0.0014	**0.0270**	-0.0213
2	-0.0083	-0.0221	**0.0501** *

Table 1. Summary of correlated character statistics for intron number and edit site number. Three independent analyses were carried out using SIMMAP 1.5. The independent analyses differ by the number of "prior draws" and "predictive sampling". **A** shows the results when both parameter were set to 45, **B** when set to 47 and **C** when set to 50. The analyses differed by the predicted number of samples. Intron number appears on the Y axis, and has 3 states, which corresponds to the absence of introns (0) or the actual number of introns present in the mitochondrial *cox2* gene of plants (1 or 2). Edit site number appears on the X axis and were coded into 3 states (0,1,2). State 0 corresponds to 0–6 edit sites, state 1 to 7–11 edit sites and state 2 to 12 or more edit sites. Correlations appear on a gradient from red (lower correlation) to green (higher correlation). *Significant positive correlation is shown in a yellow box.

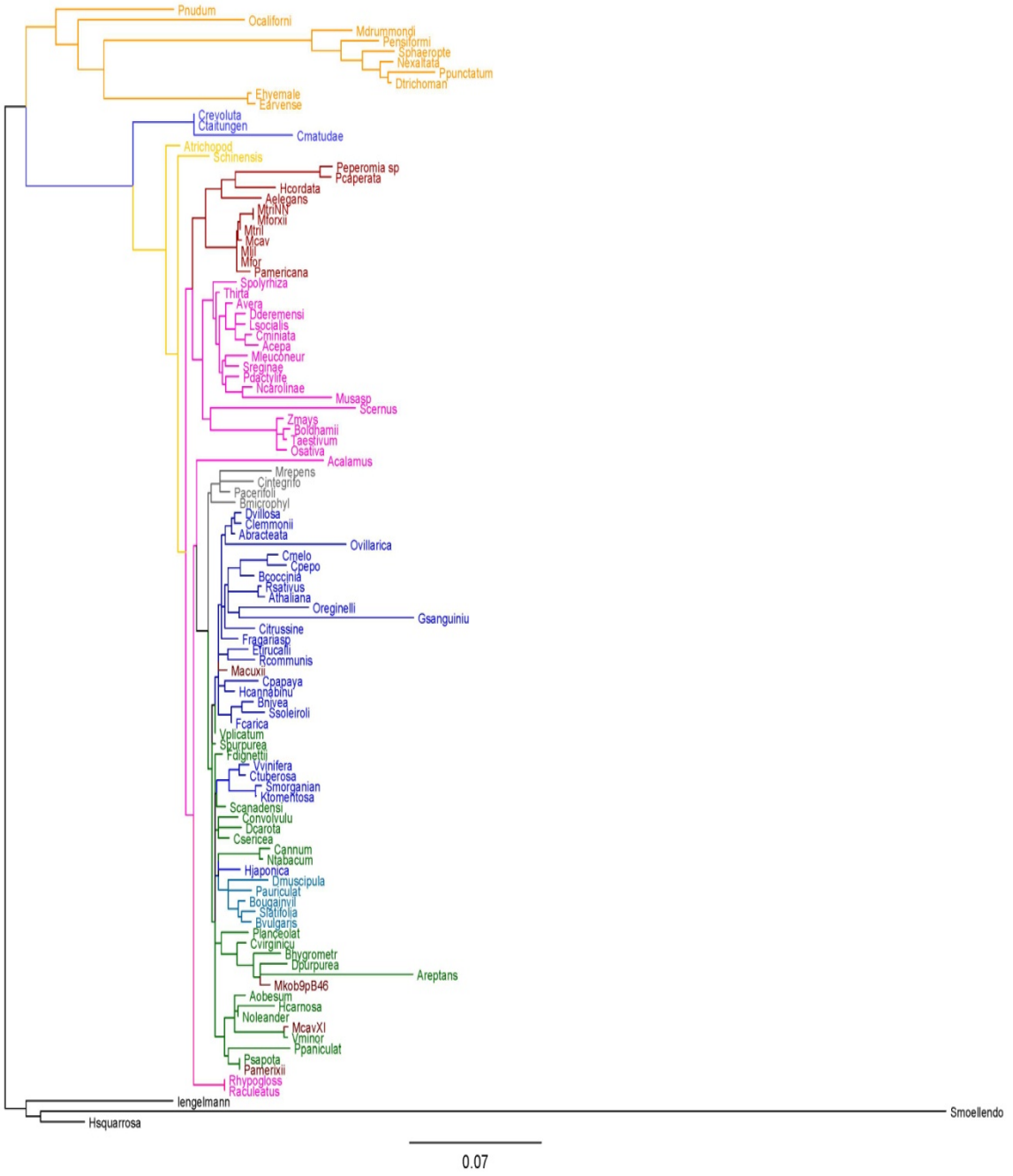


Figure 4. Maximum likelihood phylogenetic tree resulting from the analysis of 107 *cox2* cDNA sequences. Bootstrap values from 1000 replicates are shown in supplemental Figure S4. Major groups are color coded as described in Fig. 1.

DISCUSSION

Is there evidence supporting horizontal intron gain rather than loss?

Horizontal gene transfer is surprisingly common in plant mitochondria, perhaps more so in angiosperms. Twenty years of extensive research and many publications attest to this fact (Mower et al. in press). Coincidentally, the horizontally acquired group I intron located within the mitochondrial *cox1* gene has a patchy distribution among angiosperms, similar to the observed distribution in *cox2.* Taking these facts into consideration, we evaluated whether the observed intron distribution in the *cox2* gene (Fig S0) could be better explained by horizontal gain rather than intron loss.

The most comprehensive phylogenetic analysis of *cox2* i1 (Fig 1) successfully recovered a tree that is consistent with current plant taxonomy. The only species that consistently grouped in an anomalous position with moderate bootstrap support (69%-74%) was *Magnolia grandiflora* (a magnoliid), which strongly suggested a case of intron horizontal transfer (Fig S1). Unfortunately, this magnolia sequence was obtained from Genbank [accession X78418], and our degenerate primers failed to amplify the *cox2* gene or any introns for this particular species. At this point, we are unable to exclude the possibility it of being a misidentified sample or to confirm a case of intron transfer. However, considering the fact that we did not detect another probable case of horizontal intron gain in our study, it seems unlikely that *M. grandiflora's cox2* i1 is active and able to spread within or across other organellar genomes. All in all, the evidence suggests that *cox2* i1 has been vertically transferred.

Similar results were obtained from the phylogenetic analysis of *cox2* i2, for which we successfully recovered a tree that is for the most part congruent with established organismal relationships (Fig 2). The only sequence that failed to group in the expected

position was *Chionanthus virginicus* (an asterid), which was nested within the rosids, albeit with low bootstrap support (Fig S2). Our preliminary analyses, however, clearly indicate that this intron was vertically acquired. It is likely that during the final analysis, part of the *C. virginicus's* sequence was misaligned or removed by Gblocks, resulting in a lower phylogenetic signal. Further analysis is needed to examine the cause of these conflicting phylogenetic results.

Based on the presented evidence, it is clear that both introns have been vertically inherited in most if not all species. We can conclude that the observed intron distribution in the *cox2* gene of plants is more consistent with several intron losses rather than multiple horizontal gains.

Is exonization the mechanism of intron loss?

Exonization of an intron occurs when the intron sequence is no longer spliced out of the mature transcript but rather becomes part of the coding sequence.

We tested for intron loss via exonization by analyzing 46 RNA (cDNA) sequences and found no evidence of a retained intron sequence in a mature transcript. Based on the transcriptional analyses, all introns were properly spliced. From these results, we can safely exclude exonization as the mechanism of intron loss. More generally, the large size of group II introns in plants, which in *cox2* range from ~500 to >4000 bp, argues strongly against exonization. It would be highly unlikely that such large introns could be successfully translated without introducing frame shifts or premature stop codons into the coding sequence.

Is random genomic deletion the mechanism of intron loss?

In general, chromosomal deletions are not expected to precisely remove an intron. Instead, intron loss via random genomic deletion is expected to be a “messy loss”, in which part of the exon sequence is also deleted or part of the intron gets left behind. To test for random genomic deletion, the DNA sequences from 140 species that have lost either intron 1 or 2 were aligned and analyzed at the intron-exon junctions. In all cases, the observed losses were clean (data not shown), arguing against genomic deletion as the mechanism of intron loss from the *cox2* gene.

On the other hand, if selection was sufficiently strong such that imprecise genomic deletions of introns were not viable, then the only deletions we would find were precise intron removals. However, plant mitochondrial sequence alignments are often found to contain short insertions or deletions in the coding sequences, indicating that small coding changes are tolerated. Thus, it does not appear that there is such strong selection against codon insertions or removals in plant mitochondrial genes. Therefore, if genomic deletion was the mechanism of intron loss, we should expect, at least occasionally, that some of the losses would be imprecise. Because we don’t find any messy losses, we can disregard random genomic deletion as the loss mechanism.

Is retroprocessing the mechanism of intron loss?

As discussed in Chapters 1 and 2, retroprocessed genes arise from the reverse transcription of a fully spliced mRNA that undergoes homologous recombination with the native gene. In plant mitochondrial genomes, intron loss via retroprocessing is characterized by two things: a precise intron loss and a significant reduction or lack of edit sites. We tested for retroprocessing using SIMMAP 1.5. Our expectation was to

observe a strong and significant correlation between intron loss and editing loss. As shown in Table 1, the statistical analyses did show some positive correlation between intron counts and editing counts (note the green boxes for the 0 vs. 0, 1 vs. 1, and 2 vs. 2 values), but none of these values had a significant correlation that supports retroprocessing. We did, however, find a few unambiguous cases that are consistent with the retroprocessing model (e.g., *Vinca minor* and *Phlox paniculata*), which lack both introns and have lost most or all of their edit sites. But overall, the evidence for retroprocessing is rather weak.

Because several magnoliids have multiple copies of the cox2 gene that differ in intron content, one possibility is that the intronless magnoliid paralog sequences are "processed paralogs" resulting from retroprocessing. Although this sounds like a very plausible alternative explanation, it is rather unlikely because our phylogenetic analysis did not place them with other magnoliid sequences, which is the expected position for a retroprocessed gene. Furthermore, our tree was based on RNA sequences, in which all the edit sites have been converted from cytosines into uracils (which appear as thymines in cDNA). Thus, by using only cDNA sequences, we were able to successfully eliminate RNA editing sites, allowing the phylogenetic analysis to group species based on the remaining sequence variation (Bowe 1996). The unexpected phylogenetic positions of the paralogs argue against an origin by retroprocessing.

Why are the Magnolia paralogs difficult to amplify?

In general, the PCR bands for the amplified intronless magnoliid paralogs were faint, and often the PCR reactions did not amplify any band (an exception was *Magnolia tripetala*'s paralog, which always displayed a strong PCR band). The weak and inconsistent amplification results indicated that the PCR reaction was not working well for these

intronless sequences. There are several possibilities for the poor PCR results: 1) the DNA samples or PCR reagents might have a low level of DNA contamination from one or more species with intronless *cox2* genes, 2) the intronless paralogs are located in the nuclear genome, which is at a much lower copy number in a plant cell compared to the mitochondrial genome, or 3) the intronless paralogs have divergent sequences due to loss of function or to the presence of chimeric or xenologous signals.

To test for contamination, special primers were designed a few hundred base pairs inside the flanking exons to aid in the amplification process. Several successful amplifications of the paralogs were obtained with different primer combinations and all the negative controls appear negative, indicating that the faint PCR results were not due to sample contamination. Some magnolia samples were re-extracted several times and the results were consistent with the previous statements, further arguing against contamination.

We can also minimize the possibility that we are dealing with nuclear-encoded paralogs. Nuclear-encoded fragments are expected to display at least two things: a higher rate of sequence evolution and a higher affinity towards their putative progenitors (Bowe 1996; Mower et al. 2010). Our phylogenetic analyses argue against the previous statements. None of the five *Magnolia* paralogs show an increased substitution rate, and only one has a phylogenetic affinity towards other magnoliid sequences. The one exception is *Magnolia fordiana*'s paralog, labeled as "Mforxii," which does group with magnoliids (Fig. S3). However, this sequence does not display a higher substitution rate, suggesting it still resides in the mitochondrial genome. These facts argue against a case of a processed paralog that is residing in the nucleus. Further efforts using qPCR and/or

whole genome sequencing can provide more definitive evidence of the genomic location of these magnoliid paralogs.

All in all, the presented evidence suggests that the Magnolia paralogs (or xenologs) are coexisting with the native copies in the mitochondrial genome. The fact that the putative paralogs are very faint copies that required special primers points towards a degraded xenolog or chimeric copy, which we know are common in plant mitochondria (Richardson, Palmer 2007; Hao et al. 2010; Mower et al. 2010; Hepburn, Schmidt, Mower 2012).

Is HGT-GC the mechanism of intron loss?

Intron loss via HGT-GC is expected to occur through the horizontal transfer of an "intronless" xenolog, which undergoes gene conversion recombination with the native intron-containing gene resulting in a chimeric gene and a precise intron loss. It may also result in a lack of phylogenetic resolution because the chimeric gene would have conflicting signals from the native and foreign parts of the gene.

Strongly supported conflicts were found in the exon phylogeny. For instance, *Magnolia acuminata's paralog*, labeled as Macuxii, shows a higher affinity towards Fabales (core eudicots). Coincidentally, both sequences have lost both introns and have similar numbers of RNA edit sites (Table S1 and Fig S3). Similarly, *Magnolia cavaleriei's* paralog (McavXI) appears nested within a small asterid clade and it groups with 100% bootstrap support with *Vinca minor*. Interestingly, the number of RNA edit sites is strikingly similar for the two species, with 3 and 2 edit sites, respectively, and both sequences are intronless. *Magnolia kobus'* paralog (Mkob9pB46) is nested within the Lamiales with decent bootstrap support (82%) and all of the sampled Lamiales have lost

i1, with the exception of *Ajuga reptans*, which has lost both introns. Once again, the numbers of edit sites were consistent. Finally, the intronless paralog corresponding to *Persea americana* (Pamerixii) shows a strong phylogenetic affinity towards *Pouteria sapota* (91% bootstrap support) which is an asterid belonging to the order Ericales. Although most Ericales have lost their first intron, there was variation in intron content and thus, in edit site numbers. The number of RNA edit sites for this paralog was most similar to the species that have lost intron 1.The unexpected phylogenetic positions for all four paralogs strongly suggest four horizontal transfer events into the four magnoliids from the species to which they are affiliated in the tree. This is further supported by the overall similarity in intron number and edit site number in the paralogs compared to their putative donor sequences.

Furthermore, closer examination of the intron distribution in monocots (which appears as a paraphyletic group in the exon phylogeny) indicates a recent case of intron loss in Asparagales. Whereas most Asparagales have the first intron present, *Ruscus aculeatus and Ruscus hypoglossum* lack this intron. By looking at the *cox2*i1 distribution it can be inferred as a clear case of intron loss (Table S1 & Figure S0). A similar spatial pattern was observed in Acorales, but for *cox2*i2. For this particular case, it was not clear if the phylogenetic incongruence arose from an unidentified process related to the mechanism of intron loss, or if the higher substitution rates caused the anomaly.

Overall, the relationships among asterids and rosids were poorly resolved. One possible explanation for the low bootstrap support in the ML tree is the fact that we used a mitochondrial gene. Mitochondrial genes have a low level of mutation rates when compared to nuclear and plastid genes (Wolfe, Li, Sharp 1987). Perhaps an alternative topology test, such as the Shimodaira-Hasegawa (SH) test (Shimodaira, Hasegawa

1999), could determine if there is significant phylogenetic signal to rule out vertical transfer. In theory, a test like this should be able to compare our current exon ML tree with another tree that recovers “true organismal relationships” and see if significant differences exist. Moreover, it would be important to determine if the low phylogenetic signal in rosids and asterids is due to a stringent alignment or if the model we chose was not the appropriate one. Modeltest should also be able to determine the optimal model of evolution (Posada, Crandall 1998). Perhaps rosids and asterids have low sequence variability and GTR is not the most appropriate model.

Another explanation for the lack of phylogenetic resolution in the gene tree is HGT-GC. As we have seen in the previous chapter, HGT-GC was found to play an important role in the loss of two introns in *Magnolia tripetala*, and the putative *Magnolia* paralogs show many similar signs. They do not recapitulate organismal relationships (unlike the intron-containing versions), they display strongly supported conflicts, they were excluded from their vertical positions, they required downstream primers for a successful amplification, they did not show higher rate of evolution, and they grouped with intronless species with similar numbers of edit sites. Those facts suggest that the putative paralogs might be xenologs that arose by HGT (Hao et al. 2010; Mower et al. 2010).

CONCLUSIONS

In conclusion, the overall evidence points away from retroprocessing as the sole mechanism of intron loss and more towards several mechanisms of intron loss, possibly involving HGT-GC. Case in point, the observed incongruence in the exon phylogenies cannot be attributed to retroprocessing alone (since edit sites were excluded, by using a cDNA tree). Further investigation is required in *cox2* and other plant mitochondrial genes to fully uncover the role and extent of HGT-GC as a mechanism of intron loss.

ACKNOWLEDGEMENTS

We thank (Aaron Richardson, Greg Young, and Jeff Palmer) for providing Amborella cox2 sequences prior to publication; Heather Achterberg for preparing voucher specimens; Amy Hilske and Samantha Link for taking care of the plants at the Beadle center; Alex Kolbo for helping with Figure S0. This work was supported by start-up funds (JPM) and a UCARE award (DWS) from the University of Nebraska-Lincoln and by the National Science Foundation awards IOS-1027529 and MCB-1125386 (JPM).

AUTHOR'S CONTRIBUTIONS

N.J. Hepburn generated most of the *cox2* data, performed the research, and contributed with most of the figures, all under the supervision of J.P. Mower. D.W. Schmidt performed all of the RNA extractions, amplified all of the cDNA sequences, aided with some PCR amplifications and contributed with Figure S0 under the supervision of N.J. Hepburn and J.P Mower.J.P. Mower conceived the study and contributed with data analysis and interpretations.

REFERENCES

Barkman, TJ, JR McNeal, SH Lim, G Coat, HB Croom, ND Young, C W dePamphilis. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. BMC Evolutionary Biology 7:248.

Bergthorsson, U, KL Adams, B Thomason, JD Palmer. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants.

Bergthorsson, U, AO Richardson, GJ Young, LR Goertzen, JD Palmer. 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm Amborella. Proceedings of the National Academy of Sciences of the United States of America 101:17747.

Bollback, J. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. Bmc Bioinformatics 7:88.

Bonen, L. 2008. Cis-and trans-splicing of group II introns in plant mitochondria. Mitochondrion 8:26-34.

Cho, Y, YL Qiu, P Kuhlman, JD Palmer. 1998. Explosive invasion of plant mitochondria by a group I intron. Proceedings of the National Academy of Sciences 95:14244.

Davis, CC, WR Anderson, KJ Wurdack. 2005. Gene transfer from a parasitic flowering plant to a fern. Proceedings of the Royal Society B: Biological Sciences 272:2237-2242.

Davis, CC, KJ Wurdack. 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. Science 305:676-678.

Edgar, RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32:1792-1797.

Guindon, S, JF Dufayard, V Lefort, M Anisimova, W Hordijk, O Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59:307-321.

Hall, TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. p. 95-98.

Hao, W, JD Palmer. 2009. Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. Proceedings of the National Academy of Sciences 106:16728-16733.

Hao, W, AO Richardson, Y Zheng, JD Palmer. 2010. Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. Proceedings of the National Academy of Sciences 107:21576-21581.

Hepburn, NJ, DW Schmidt, JP Mower. 2012. Loss of two introns from the Magnolia tripetala mitochondrial cox2 gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. Molecular Biology and Evolution.

Huelsenbeck, JP, R Nielsen, JP Bollback. 2003. Stochastic mapping of morphological characters. Systematic Biology 52:131-158.

Joly, S, L Brouillet, A Bruneau. 2001. Phylogenetic implications of the multiple losses of the mitochondrial coxII. i3 intron in the angiosperms. International Journal of Plant Sciences 162:359-373.

Keeling, PJ, JD Palmer. 2008. Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics 9:605-618.

Kudla, J, F Albertazzi, D Blazević, M Hermann, R Bock. 2002. Loss of the mitochondrial cox2 intron 1 in a family of monocotyledonous plants and utilization of

mitochondrial intron sequences for the construction of a nuclear intron. Molecular Genetics and Genomics 267:223-230.
Lambowitz, AM, S Zimmerly. 2011. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harbor Perspectives in Biology 3.
Mower, JP. 2009. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic acids research 37:W253-W259.
Mower, JP, Stefanovi, S cacute, GJ Young, JD Palmer. 2004. Plant genetics: gene transfer from parasitic to host plants. Nature 432:165-166.
Mower, JP, S Stefanović, W Hao, JS Gummow, K Jain, D Ahmed, JD Palmer. 2010. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. BMC biology 8:150.
Nickrent, DL, RJ Duff, AE Colwell, AD Wolfe, N Young, K Steiner, C DePamphilis. 1998. Molecular phylogenetic and evolutionary studies of parasitic plants. Molecular systematics of plants II: DNA sequencing:211-241.
Richardson, AO, JD Palmer. 2007. Horizontal gene transfer in plants. Journal of experimental botany 58:1-9.
Ronquist, F. 2004. Bayesian inference of character evolution. Trends in Ecology & Evolution 19:475-481.
Sanchez-Puerta, MV, CC Abbona, S Zhuo, EJ Tepe, L Bohs, RG Olmstead, JD Palmer. 2011. Multiple recent horizontal transfers of the cox1 intron in Solanaceae and extended co-conversion of flanking exons. BMC evolutionary biology 11:277.
Sanchez-Puerta, MV, Y Cho, JP Mower, AJ Alverson, JD Palmer. 2008. Frequent, phylogenetically local horizontal transfer of the cox1 group I intron in flowering plant mitochondria. Molecular biology and evolution 25:1762-1777.
Talavera, G, J Castresana. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56:564-577.
Tamura, K, D Peterson, N Peterson, G Stecher, M Nei, S Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28:2731-2739.
Vaughn, JC, MT Mason, GL Sper-Whitis, P Kuhlman, JD Palmer. 1995. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric coxI gene of Peperomia. Journal of molecular evolution 41:563-572.
Woloszynska, M, T Bocer, P Mackiewicz, H Janska. 2004. A fragment of chloroplast DNA was transferred horizontally, probably from non-eudicots, to mitochondrial genome of Phaseolus. Plant molecular biology 56:811-820.
Won, H, SS Renner. 2003. Horizontal gene transfer from flowering plants to Gnetum. Proceedings of the National Academy of Sciences 100:10824.

SUPPLEMENTARY TABLES

Supplementary Table S1

Group	Order	Family	Species	Source	Source #	i94	i1(i373)	i2(i691)	i_#	# EditSites
Lycopodiophyta	Isoetales	Isoetaceae	Isoetes engelmannii	Genbank	FJ390841,HQ616427	+	-	-	0	14
Lycopodiophyta	Lycopodiales	Lycopodiaceae	Huperzia lucidula	Genbank	DQ677486		+	+	2	-
Lycopodiophyta	Lycopodiales	Lycopodiaceae	Huperzia squarrosa	Genbank	JQ002659	+	+	+	3	4
Lycopodiophyta	Selaginellales	Selaginellaceae	Selaginella moellendorffii	Genbank	JF338144, JF338146,JF276244	+	+	+	3	96
Monilophytes	Equisetales	Equisetaceae	Equisetum arvense	Genbank	FJ376599		+	-	1	1
Monilophytes	Equisetales	Equisetaceae	Equisetum hyemale	This study	1075		+	-	1	0
Monilophytes	Cyatheales	Cyatheaceae	Sphaeropteris sp	This study	1104		+	-	1	17
Monilophytes	Polypodiales	Davalliaceae	Davallia trichomanoides	This study	1108		+	-	1	16
Monilophytes	Polypodiales	Dryopteridaceae	Dryopteris arguta	This study	1076		+	?	?	-
Monilophytes	Polypodiales	Nephrolepidacea	Nephrolepis exaltata	This study	1133		+	-	1	15
Monilophytes	Polypodiales	Polypodiaceae	Polypodium punctatum	This study	1077		+	-	1	36
Monilophytes	Polypodiales	Pteridaceae	Pteris ensiformis	This study	N/A		+	-	1	13
Monilophytes	Salviniales	Marsileaceae	Marsilea drummondii	This study	1078		+	-	1	29
Monilophytes	Ophioglossales	Ophioglossaceae	Ophioglossum californicum	This study	1060		-	-	0	10
Monilophytes	Polypodiales	Aspleniaceae	Asplenium nidus	Genbank	DQ677483		+	?	?	-
Monilophytes	Psilotales	Psilotaceae	Psilotum nudum	This study	1074		-	-	0	10
Gymnosperms	Coniferales	Cupressaceae	Cupressus lusitanica	This study	N/A		-?	-	0?	-
Gymnosperms	Coniferales	Cupressaceae	Metasequoia glyptostroboides	This study	N/A		+	-	1	-
Gymnosperms	Coniferales	Cupressaceae	Taxodium distichum	This study	N/A		+?	-	1?	-
Gymnosperms	Coniferales	Podocarpaceae	Podocarpus macrophyllus	This study	1105		+?	-	1?	-
Gymnosperms	Cycadales	Cycadaceae	Cycas revoluta	This study	1102		+	+	2	22
Gymnosperms	Cycadales	Cycadaceae	Cycas taitungensis	Genbank	AP009381		+	+	2	27
Gymnosperms	Cycadales	Zamiaceae	Ceratozamia matudae	This study	N/A		+	+	2	24
Gymnosperms	Cycadales	Zamiaceae	Dioon edule	This study	1110		+	+	2	22
Gymnosperms	Cycadales	Zamiaceae	Zamia floridana	This study	1101		+	+	2	21
Gymnosperms	Cycadales	Zamiaceae	Zamia pumila	This study	N/A		+	+	2	-
Gymnosperms	Ginkgoales	Ginkgoaceae	Ginkgo biloba	This study	N/A		+	+	2	19
Basal Angiosperms	Amborellales	Amborellaceae	Amborella trichopoda	N/A	N/A		+	+	2	21
Basal Angiosperms	Austrobaileyale	Schisandraceae	Schisandra chinensis	Genbank	JQ317136		+	+	2	20
Magnoliids	Laurales	Calycanthaceae	Calycanthus floridus	Genbank	JQ317141		+	+	2	16
Magnoliids	Laurales	Lauraceae	Laurus nobilis	Genbank	AY832094, AY832104		+	?	?	13
Magnoliids	Laurales	Lauraceae	Persea americana	Genbank	JQ317143		+	+	2	16
Magnoliids	Laurales	Lauraceae	Persea americana xii	This study	1112		-	-	0	11
Magnoliids	Magnoliales	Annonaceae	Asimina triloba	Genbank	JQ317142		+	+	2	16
Magnoliids	Magnoliales	Magnoliaceae	Liriodendron tulipifera	This study	1001		+	+	2	18
Magnoliids	Magnoliales	Magnoliaceae	Magnolia acuminata	Genbank	JQ317147		+	+	2	16
Magnoliids	Magnoliales	Magnoliaceae	Magnolia acuminata paralog xii	This study	1007		-	-	0	10
Magnoliids	Magnoliales	Magnoliaceae	Magnolia cavaleriei paralog xi	This study	2012		-	-	0	3
Magnoliids	Magnoliales	Magnoliaceae	Magnolia cavaleriei var. platypeta	Genbank	JQ317148		+	+	2	15
Magnoliids	Magnoliales	Magnoliaceae	Magnolia fordiana	Genbank	JQ317146		+	+	2	16
Magnoliids	Magnoliales	Magnoliaceae	Magnolia fordiana paralog xii	This study	2004		-	-	0	10
Magnoliids	Magnoliales	Magnoliaceae	Magnolia grandiflora	Genbank	X78418		+	?	?	
Magnoliids	Magnoliales	Magnoliaceae	Magnolia kobus	Genbank	JQ317149		+	+	2	16
Magnoliids	Magnoliales	Magnoliaceae	Magnolia kobus 9pB46	This study	1009		-	-	0	9
Magnoliids	Magnoliales	Magnoliaceae	Magnolia liliifera	Genbank	JQ317145		+	+	2	16
Magnoliids	Magnoliales	Magnoliaceae	Magnolia pyramidata	Genbank	JQ317144		+	+	2	16
Magnoliids	Magnoliales	Magnoliaceae	Magnolia stellata	Genbank	JQ317150		+	+	2	16
Magnoliids	Magnoliales	Magnoliaceae	Magnolia tamaulipana	Genbank	JQ317151		+	+	2	16
Magnoliids	Magnoliales	Magnoliaceae	Magnolia tripetala I	Genbank	JQ317152		+	+	2	17
Magnoliids	Magnoliales	Magnoliaceae	Magnolia tripetala N	Genbank	JQ317133		-	-	0	14
Magnoliids	Magnoliales	Magnoliaceae	Magnolia tripetala NNN	Genbank	JQ317132		-	-	0	14
Magnoliids	Magnoliales	Magnoliaceae	Magnolia yuyuanensis	This study	2009		+	+	2	?
Magnoliids	Magnoliales	Magnoliaceae	Magnolia yuyuanensis paralog iii	This study	2009		-	-	0	0
Magnoliids	Piperales	Aristolochiaceae	Asarum sp	Genbank	AY832096, AY832105		+	?	?	10
Magnoliids	Piperales	Aristolochiacee	Aristolochia elegans	Genbank	JQ317137		+	+	2	16
Magnoliids	Piperales	Piperaceae	Peperomia caperata	Genbank	JQ317135		+	+	2	18
Magnoliids	Piperales	Piperaceae	Peperomia spp	This study	N/A		+	+	2	-
Magnoliids	Piperales	Piperaceae	Piper betle	Genbank	AY832091, AY832102		+	+	2	16
Magnoliids	Piperales	Saururaceae	Houttuynia cordata	Genbank	JQ317139		+	+	2	14
Monocots	Acorales	Acoraceae	Acorus calamus	Genbank	AJ006146		+	-	1	8
Monocots	Acorales	Acoraceae	Acorus gramineus	This study	1144		+	-	1	-
Monocots	Alismatales	Araceae	Spirodela polyrhiza	Genbank	JQ804980		+	-	1	17
Monocots	Arecales	Arecaceae	Phoenix dactylifera	Genbank	JN375330		+	+	2	17
Monocots	Asparagales	Agavaceae	Agave americana	This study	N/A		?	+	?	-
Monocots	Asparagales	Alliaceae	Allium cepa (NJGH)	This study	N/A		+	-	1	-
Monocots	Asparagales	Alliaceae	Allium cepa	Genbank	GU253305,GU253307		+	-	1	16
Monocots	Asparagales	Alliaceae	Allium schoenoprasum	This study	N/A		+	-	1	-
Monocots	Asparagales	Amaryllidaceae	Allium senescens	This study	1151		+	+	2	-
Monocots	Asparagales	Amaryllidaceae	Clivia miniata	This study	1115		+	+	2	5
Monocots	Asparagales	Asparagaceae	Agave attenuata	Genbank	AY832092,AY832103		+	?	?	14
Monocots	Asparagales	Asparagaceae	Dracaena deremensis	This study	N/A		+	+	2	17
Monocots	Asparagales	Asparagaceae	Ruscus aculeatus	This study	N/A		-	+	1	19
Monocots	Asparagales	Asparagaceae	Ruscus hypoglossum	This study	1121		-	+	1	16
Monocots	Asparagales	Asperagaceae	Asparagus densiflorus	This study	N/A		+	+	2	-
Monocots	Asparagales	Asphodelaceae	Aloe vera	This study	N/A		+	+	2	15

Monocots	Asparagales	Hyacinthaceae	Ledebouria socialis	This study	N/A	+	+	2	18
Monocots	Commelinales	Pontederiaceae	Eichhornia crassipes	Genbank	AY832093, AY832099	+	?	?	13
Monocots	Liliales	Liliaceae	Tricyrtis hirta	This study	N/A	+	+	2	17
Monocots	Poales	Bromeliaceae	Neoregelia carolinae	This study	N/A	+	+	2	17
Monocots	Poales	Cyperaceae	Scirpus cernus	This study	N/A	-	-	0	1
Monocots	Poales	Poaceae	Oryza sativa	Genbank	DQ167399	+	-	1	19
Monocots	Poales	Poaceae	Bambusa oldhamii	Genbank	EU365401	+	-	1	15
Monocots	Poales	Poaceae	Ferrocalamus rimosivaginus	Genbank	JN120789	+	-	1	15
Monocots	Poales	Poaceae	Sorghum bicolor	Genbank	DQ984518	+	-	1	15
Monocots	Poales	Poaceae	Tripsacum dactyloides	Genbank	DQ984517	+	-	1	15
Monocots	Poales	Poaceae	Zea mays	Genbank	AY506529	+	-	1	18
Monocots	Poales	Poaceae	Triticum aestivum	Genbank	AP008982/ X52867	+	-	1	16
Monocots	Zingiberales	Marantaceae	Maranta leuconeura	This study	N/A	-	-	0	16
Monocots	Zingiberales	Musaceae	Musa sp	This study	N/A	+	+	2	12
Monocots	Zingiberales	Strelitziaceae	Strelitzia reginae	This study	1141	+	+	2	17
Monocots	Zingiberales	Zingiberaceae	Hedychium sp	This study	N/A	+	+	2	17
Basal Eudicots	Buxales	Buxaceae	Buxus microphylla	Genbank	JQ317138	+	+	2	16
Basal Eudicots	Proteales	Platanaceae	Platanus x acerifolia	Genbank	JQ317140	+	+	2	19
Basal Eudicots	Ranunculales	Berberidaceae	Mahonia repens	Genbank	JQ317131	-	-	0	12
Basal Eudicots	Ranunculales	Papaveraceae	Eschscholzia californica	Genbank	EU924190	-	?	?	14
Basal eudicots	Ranunculales	Ranunculaceae	Clematis integrifolia	Genbank	JQ317134	-	+	1	12
Rosids	Brassicales	Brassicaceae	Arabidopsis thaliana	Genbank	Y08501	-	+	1	15
Rosids	Brassicales	Brassiceae	Brassica napus	Genbank	AP006444	-	+	1	13
Rosids	Brassicales	Brassiceae	Raphanus sativus	Genbank	AF036387	-	+	1	15
Rosids	Brassicales	Caricaceae	Carica papaya	This study	N/A	-	+	1	8
Rosids	Cucurbitales	Begoniaceae	Begonia coccinia	This study	N/A	-	+	1	12
Rosids	Cucurbitales	Cucurbitaceae	Citrullus lanatus	Genbank	GQ856147	-	+	1	10
Rosids	Cucurbitales	Cucurbitaceae	Cucumis melo	Genbank	JF412792	-	-	0	13
Rosids	Cucurbitales	Cucurbitaceae	Cucumis sativus	Genbank	HQ860792	-	-	0	13
Rosids	Cucurbitales	Cucurbitaceae	Cucurbita pepo	Genbank	GQ856148	-	+	1	12
Rosids	Fabales	Fabaceae	Amphicarpaea bracteata	Genbank	AF207682	-	-	0	13
Rosids	Fabales	Fabaceae	Cologania lemmonii	Genbank	AF207746	-	-	0	14
Rosids	Fabales	Fabaceae	Dumasia villosa	Genbank	AF207747	-	-	0	15
Rosids	Fabales	Fabaceae	Glycine max	Genbank	X04825	-	-	0	13
Rosids	Fabales	Fabaceae	Lespedeza formosa	Genbank	AF208162	-	-	0	13
Rosids	Fabales	Fabaceae	Lotus japonicus	Genbank	JN872551	-	-	0	13
Rosids	Fabales	Fabaceae	Millettia pinnata	Genbank	JN872550	-	-	0	13
Rosids	Fabales	Fabaceae	Pisum sativum	Genbank	AJ414385	-	-	0	13
Rosids	Fagales	Fagaceae	Fagus sylvatica	This study	1012	-	+	1	12
Rosids	Geraniales	Geraniaceae	Geranium sanguinium	Genbank	JQ317153	-	-	0	2
Rosids	Malpighiales	Euphorbiaceae	Euphorbia tirucalli	This study	N/A	-	+	1	4
Rosids	Malpighiales	Euphorbiaceae	Ricinus communis	Genbank	HQ874649	-	+	1	11
Rosids	Malpighiales	Passifloraceae	Adenia karibaensis	This study	N/A	-	+	1	-
Rosids	Malpighiales	Phyllanthaceae	Breynia disticha	This study	N/A	-	+	1	10
Rosids	Malvales	Malvaceae	Hibiscus cannabinus cultivar P3B	Genbank	HM535789	-	+	1	11
Rosids	Myrtales	Onagraceae	Oenothera villaricae	Genbank	X00212	-	-	0	15
Rosids	Oxalidales	Oxalidaceae	Oxalis reginelli	This study	N/A	-	+	1	10
Rosids	Oxalidales	Oxalidaceae	Oxalis rubra	This study	N/A	-	+	1	-
Rosids	Rosales	Moraceae	Ficus carica	This study	N/A	-	+	1	1
Rosids	Rosales	Rosaceae	Fragaria sp	This study	N/A	-	-	0	13
Rosids	Rosales	Urticaceae	Boehmeria nivea	Genbank	EU122339	-	-	0	0
Rosids	Rosales	Urticaceae	Soleirolia soleirolii	This study	N/A	-	+	1	0
Rosids	Sapindales	Rutaceae	Citrus sinensis	This study	N/A	-	+	1	3
Rosids	Saxifragales	Crassulaceae	Kalanchoe tomentosa	This study	N/A	+	+	2	12
Rosids	Saxifragales	Crassulaceae	Kalanchoe tubiflora	This study	N/A	+	+	2	11
Rosids	Saxifragales	Crassulaceae	Sedum morganianum	This study	N/A	+	+	2	12
Rosids	Saxifragales	Hamamelidaceae	Hamamelis japonica	This study	1017	-	-	0	11
Rosids	Vitales	Vitaceae	Cissus tuberosa	This study	1126	+	+	2	16
Rosids	Vitales	Vitaceae	Vitis vinifera	Genbank	FM179380	+	+	2	19
Caryophyllids	Caryophyllales	Amaranthaceae	Beta vulgaris	Genbank	BA000009	+	-	1	9
Caryophyllids	Caryophyllales	Caryophyllaceae	Silene latifolia	Genbank	HM562727	-	-	0	3
Caryophyllids	Caryophyllales	Droseraceae	Dionaea muscipula	This study	1082	+	+	2	11
Caryophyllids	Caryophyllales	Nyctaginaceae	Bougainvillea sp	This study	1120	+	+	2	8
Caryophyllids	Caryophyllales	Plumbaginaceae	Plumbago auriculata	This study	1128	-	-	0	1
Asterids	Apiales	Apiaceae	Daucus carota subsp. sativus	Genbank	X63625	+	+	2	14
Asterids	Apiales	Araliaceae	Panax ginseng	Genbank	AF033560	+	+	2	7
Asterids	Asterales	Asteraceae	Artemisia dranunculus	This study	N/A	-	?	?	-
Asterids	Cornales	Cornaceae	Cornus sericea	This study	N/A	+	+	2	13
Asterids	Cornales	Hydrangeaceae	Hydrangea macrophylla	This study	1070	-	+	1	15
Asterids	Dipsacales	Adoxaceae	Sambucus canadensis	This study	1069	+	-	1	15
Asterids	Dipsacales	Adoxaceae	Viburnum plicatum	This study	1006	-	+	1	13
Asterids	Ericales	Actinidiaceae	Actinidia sp	This study	N/A	-	+	1	14
Asterids	Ericales	Fouquieriaceae	Fouqieria dignettii	This study	N/A	-	+	1	14
Asterids	Ericales	Polemoniaceae	Phlox paniculata	This study	N/A	-	-	0	0
Asterids	Ericales	Sapotaceae	Pouteria sapota	This study	1118	-	+	1	13
Asterids	Ericales	Sarraceniacee	Sarracenia purpurea	This study	1079	-	+	1	14
Asterids	Ericales	Theaceae	Camellia sinensis	Genbank	AY845271	+	+	2	6
Asterids	Gentianales	Apocynaceae	Adenium obesum	This study	N/A	+	-	1	8
Asterids	Gentianales	Apocynaceae	Ceropegia woodii	This study	N/A	+	-	1	7
Asterids	Gentianales	Apocynaceae	Hoya carnosa	This study	N/A	+	-	1	8

Asterids	Gentianales	Apocynaceae	Nerium oleander	This study	N/A	+	-	1	7
Asterids	Gentianales	Apocynaceae	Stapelia hirsuta	This study	N/A	+	-	1	8
Asterids	Gentianales	Apocynaceae	Vinca minor	This study	1068	-	-	0	2
Asterids	Gentianales	Rubiaceae	Coffea arabica	This study	N/A	-	-	0	2
Asterids	Gentianales	Rubiaceae	Pentas lanceolata	This study	1117	+	-	1	6
Asterids	Lamiales	Gesneriaceae	Boea hygrometrica	Genbank	JN107812	-	+	1	10
Asterids	Lamiales	Lamiaceae	Ajuga reptans	This study	1062	-	-	0	10
Asterids	Lamiales	Oleaceae	Chionanthus virginicus	This study	1004	-	+	1	7
Asterids	Lamiales	Phrymaceae	Mimulus guttatus	Genbank	JN098455	-	+	1	9
Asterids	Lamiales	Plantaginaceae	Digitalis purpurea	Genbank	Unpublished	-	+	1	6
Asterids	Solanales	Convolvulaceae	Convolvulus sp	This study	1056	+	-	1	10
Asterids	Solanales	Solanaceae	Capsicum annum	This study	1127	+	-	1	9
Asterids	Solanales	Solanaceae	Nicotiana tabacum	Genbank	BA000042	+	-	1	14
Asterids	Solanales	Solanaceae	Petunia x hybrida	Genbank	X17394	+	-	1	13
Asterids	Solanales	Solanaceae	Solanum tuberosum	Genbank	DQ185064	+	-	1	12

Table S1. Taxonomy, Source and Genbank accession numbers are provided for plant materials used in this study. Abbreviations: University of Nebraska-Lincoln, Beadle Center (UNL-BC); University of Nebraska-Lincoln, Earl G. Maxwell Arboretum (UNL-EGMA) and the JC Raulstone Arboretum (JC-RA) .

Supplementary Table S2

Anchoring position	Primer name PCR:	Primer Sequences (5'-3')
EXON 1	EMBRYO-cox2e1-F88	GCAACACCTATNATGCAAGGAAT
	GYMNO-cox2e1-F155	TCGTATYAYGGATGTTGGTTCG
	MONOCOT-cox2e1-F121	CATCACGATATCTTTTTCTTCCTC
	PTERI-cox2e1-F76	GACGCAGCCAYACCTATGATG
	PTERI-cox2e1-F82	GCAGCCAYACCTATGATGYAAGG
	UNI-cox2e1-F1	TTGTGATGCWGCGGAACC
	UNI-cox2e1-F2	CCNTGGCAATTAGGATYTCAAGA
EXON 2	EMBRYO-cox2e2-F490	GTGGACAATMGAGTGGTTGTNCC
	EMBRYO-cox2e2-R613	AACGACCRGGTACAGCATC
	PTERI-cox2e2-R1584	TAAACTCCCTCTCATTATGTA
	UNI-cox2e2-F3	CGTTTATTAGAAGTNGACAATMGAGT
	UNI-cox2e2-R3	ACCTRAGGAAGGTACAGCC
EXON 3	UNI-cox2e3-R1	GAGGATTAATTGATTGRATACCCR
	UNI-cox2e3-R2	CTA GRA ACA GCT TCT ACG ACG AT

Table S2. Primers used in this study, anchoring in exon 1, 2 or 3 of the *cox2* gene of plants.

Supplementary Table S3

Anchoring position INTRON 1(i373)	**Primer name** PCR:	**Primer Sequences (5'-3')**
Degenerate primers	ANG-cox2i1-F1133	CTG GAA CGT GGG AAT TCG
	ANG-cox2i1-F1348	CTA CAA CTT CRC CGA GCC
	ANG-cox2i1-R1150	CGA ATT CCC ACG TTC CAG
	ANG-cox2i1-R1365	GGC TCG GYG AAG TTG TAG
	EMB-cox2i1-F1672	GCG TGG ARA GCT KTT TGC
	EMB-cox2i1-F513	GCC AGA AAC GGG RAG TTG
	EMB-cox2i1-F893	TKA CCG AAR GGG ACC AGC
	EMB-cox2i1-R1689	GCA AAM AGC TYT CCA CGC
	EMB-cox2i1-R530	CAA CTY CCC GTT TCT GGC
	EMB-cox2i1-R912	GCT GGT CCC YTT CGG TMA
	SPE-cox2i1-F1431	ATT NGC TTT CCG TGG TGA ACT
	SPE-cox2i1-R1451	AGT TCA CCA CGG AAA GCN AAT
Specific	Asplenium-cox2i1-F1507	CTGTCAGGGACGAGAGAATAGTACGAG
	Asplenium-cox2i1-R1534	CTCGTACTATTCTCTCGTCCCTGACAG
	Aurahetero-cox2i1-F117	GGAGGCTCTCTACTGCTATAATCG
	Aurahetero-cox2i1-F1307	GTGCTCGTTATATGGATCTCGC
	Aurahetero-cox2i1-F2045	TACGACTGGACAACTACCGATG
	Aurahetero-cox2i1-F2461	CGAGTACATAAATGACCTGCC
	Aurahetero-cox2i1-F658	CGGCATTCTCAATAGTAGCG
	Aurahetero-cox2i1-F838	GTTCTCCTTCCATCCACCTTC
	Aurahetero-cox2i1-R1328	GCGAGATCCATATAACGAGCAC
	Aurahetero-cox2i1-R141	CGATTATAGCAGTAGAGAGCCTCC
	Aurahetero-cox2i1-R1635	GCAGCGAGCAAAAGTATGGTG
	Aurahetero-cox2i1-R2076	CAGGCAAGTATTCGTCTATTGGC
	Aurahetero-cox2i1-R684	CGCTACTATTGAGAATGCCG
	Aurahetero-cox2i1-R858	GAAGGTGGATGGAAGGAGAAC
	Buxus-cox2i1-F1496	CGCTGGAATGAGCAGTAGTAGGATC
	Calycanflo-cox2i1-F1400	GCCTACACTTGCTCCGTACCTTACC
	Calycanflo-cox2i1-R1120	TGACAAAGGATAGACTAGACTGGAC
	Ceratozamia-cox2i1-F2037	GAGAGGGATTTAGAGAGAATAGCA
	Ceratozamia-cox2i1-F2247	GCTATGTTCTAAATGAGAGGG
	Ceratozamia-cox2i1-F2370	CTAAATGAGAGTTGGAAACTTAGAG
	Ceratozamia-cox2i1-R2106	CTAGAGATTGTTCAGCCTCC
	Ceratozamia-cox2i1-R2280	GTTCCAACTCTCATTTAGAACATAGC
	Cissus-cox2i1-F1220	CTCAATCCATAGCGGTCTCAC
	Cissus-cox2i1-R1240	GTGAGACCGCTATGGATTGAG
	Cycas-cox2i1-F1883	GCCATAGCGATTAGATTCTTG
	Cycas-cox2i1-R1599	GATAGTGTCCAAGCATTGATTAG
	Cycas-cox2i1-R2210	GGTTGTCCTATACCGTGTGAC
	Equisetum-cox2i1-F1275	GTCAACCGCCGTTCAGGA
	Equisetum-cox2i1-R1292	TCCTGAACGGCGGTTGAC
	Ginkgo-cox2i1-F1334	CCAGTCCATCCGACCATTCA
	Ginkgo-cox2i1-F1853	GATAGAACCAGAATCACCCAC
	Ginkgo-cox2i1-F195	CGGAATGGGCAAGCACTTTC
	Ginkgo-cox2i1-R1353	TGAATGGTCGGATGGACTGG
	Ginkgo-cox2i1-R1877	GTGGGTGATTCTGGTTCTATC

Ginkgo-cox2i1-R2145	GAAAGTGCTTGCCCATTCCG
Houttuyniac-cox2i1-1R1286	GAAGCGAGAAGCAGAGCCT
Houttuyniac-cox2i1-F1268	AGGCTCTGCTTCTCGCTTC
Magnoliaacu-cox2i1-F1492	GCTGCTAAAATGGAAGGGCG
Magnoliaacu-cox2i1-F195	GATAGACTTGTACTAGAAACTAGCG
Magnoliaacu-cox2i1-R219	CGCTAGTTTCTAGTACAAGTCTATC
Magnoliako-cox2i1-F1718	CGTAATAGGGGCTTTTCAGC
Magnoliako-cox2i1-R1737	GCTGAAAAGCCCCTATTACG
Magnoliaste-cox2i1-F1500	GCACCATTACTATTGACGGG
Magnoliaste-cox2i1-F581	CGACCAGTTCACCACGGAAAGCG
Magnoliaste-cox2i1-R1519	CCCGTCAATAGTAATGGTGC
Magnoliaste-cox2i1-R604	CGCTTTCCGTGGTGAACTGGTCG
Magnoliaste-cox2i1-R864	CACCGTTCACGACATCCCTTGC
Magnoliatri-cox2i1-F556	GTGTGGAGCGATATACGAGAAATAGA
Magnoliatri-cox2i1-R531	TCTATTTCTCGTATATCGCTCCACAC
Musa-cox2i1-R871	CTCCTCGGTTTCGTAGTAGAAG
Peperomia-cox2i1-F1469	GTTTCTATGCTTATTGTGACTTACG
Peperomia-cox2i1-R1493	CGTAAGTCACAATAAGCATAGAAAC
Persea-cox2i1-F1461	TACACTTGCTCCTTCGTTTGCTGG
Persea-cox2i1-R1484	CCAGCAAACGAAGGAGCAAGTGTA
Peteris-cox2i1-F1346	GCCGACCCAACCTTATGAGTATTC
Platanus-cox2i1-R1274	AGAGTGATTCCTCACCTATCCTGTC
Schisandra-cox2i1-F1476	AGAAACTAGCGGTGAACAACGGAG
Schisandra-cox2i1-F2103	CTTATGAGTATTCGGACTATAACAGT
Schisandra-cox2i1-R1500	CTCCGTTGTTCACCGCTAGTTTCT
Zamia-cox2i1-F1362	GTCAGGGACGAGGGAATAG
Zamia-cox2i1-F1962	CACATAGTGCGAGACTGAGACTG
Zamia-cox2i1-F1978	CTACAACTTCGCCGACTAGC
Zamia-cox2i1-F2731	TTGATTAAGATTCCCATATTGATGA
Zamia-cox2i1-R1383	CTATTCCCTCGTCCCTGAC
Zamia-cox2i1-R1984	CAGTCTCAGTCTCGCACTATGTG
Zamia-cox2i1-R1997	GCTAGTCGGCGAAGTTGTAG
Zamia-cox2i1-R2757	TCATCAATATGGGAATCTTAATCAA
Zamiaflo-cox2i1-F2225	TGCTATGTCCTGAATGAGAG
Zamiaflo-cox2i1-R1586	GTTGGACAGATGCCACTGTGC
Zamiaflo-cox2i1-R2120	CAACTCTCACTCAGGACATAGC

Table S3. Primers used in this study, anchoring in the *cox2* intron1 (i373) of plants.

Supplementary Table S4

Anchoring position INTRON2 (i691)	**Primer name** PCR:	**PrimerSequences(5'-3')**
Degenerate	ANG-cox2i2-F959	GAAGGACCTGCAACGGCA
	ANG-cox2i2-R977	TGCCGTTGCAGGTCCTTC
	EMB-cox2i2-F1303	AGATCTAGGAGKGTGAGCAG
	EMB-cox2i2-F540	CGAGCGAGTGGTTAGTRG
	EMB-cox2i2-F757	GGAGTTNAGAGGCCTYATAGT
	EMB-cox2i2-R1322	CTGCTCACMCTCCTAGATCT
	EMB-cox2i2-R557	CYACTAACCACTCGCTCG
	EMB-cox2i2-R777	ACTATRAGGCCTCTNAACTCC
	GYMN-cox2i2-F1270	GCCAGCCTATTGATTCATGG
	GYMN-cox2i2-F3668	ACTAAGGTTCTACGATCGATCA
	GYMN-cox2i2-R1289	CCATGAATCAATAGGCTGGC
	GYMN-cox2i2-R3689	TGATCGATCGTAGAACCTTAGT
Specific	Adeniaka-cox2i2-F1210	ATTGGGTCGTATTGAACGGGT
	Adeniaka-cox2i2-F660	CTCGCTTGCTAGATTTCGGATTC
	Adeniaka-cox2i2-F991	GATCTGGCAGTGGATTGCTATTGAC
	Adeniaka-cox2i2-R1015	GTCAATAGCAATCCACTGCCAGATC
	Adeniaka-cox2i2-R1230	ACCCGTTCAATACGACCCAAT
	Adeniaka-cox2i2-R683	GAATCCGAAATCTAGCAAGCGAG
	Aristoele-cox2i2-F1613	GAAACGCTCAGGAAAGGAG
	Aristoele-cox2i2-F1731	CTACGAAGCAAGTGAGCGGAG
	Aristoele-cox2i2-R1630	CTCCTTTCCTGAGCGTTTC
	Asiminatri-cox2i2-F2779	GACTTATCCATTCTTGCGATTCCC
	Asiminatri-cox2i2-F3214	GTGTGAAACGCTCAGGAAAGGAG
	Asiminatri-cox2i2-R2755	GGGAATCGCAAGAATGGATAAGTC
	Asiminatri-cox2i2-R3192	CTCCTTTCCTGAGCGTTTCACAC
	Asplenium-cox2i2-F1155	CGGGCAAGATAGATGGAAAGGAGACG
	Asplenium-cox2i2-F560	GAAGTAGGGAGTCTCGGTTCGCTG
	Asplenium-cox2i2-R1180	CGTCTCCTTTCCATCTATCTTGCCCG
	Asplenium-cox2i2-R580	CAGCGAACCGAGACTCCCTACTTC
	Begonia-cox2i2-F1452	ATTGCTCCGCTTATAGAATAGAATG
	Begonia-cox2i2-R1476	CATTCTATTCTATAAGCGGAGCAAT
	Bergenia-cox2i2-F1172	GACTTTTCTTGCGATTCCCA
	Bergenia-cox2i2-R1191	TGGGAATCGCAAGAAAAGTC
	Calycanflo-cox2i2-F1802	GCGGCGGACTTCCAACACCT
	Calycanflo-cox2i2-F4515	GGTGAGATAAACGGAAGAGTAGC
	Calycanflo-cox2i2-R1821	AGGTGTTGGAAGTCCGCCGC
	Cissus-cox2i2-F1513	CCTTGTATAGGTTGGAGCTATG
	Cissus-cox2i2-R1534	CATAGCTCCAACCTATACAAGG
	Dionaea-cox2i2-F1015	TCCTCAAAGTCCTCAATGAACT
	Dionaea-cox2i2-R1036	AGTTCATTGAGGACTTTGAGGA
	Dioon-cox2i2-F1827	GGTGAGGCACTGTACTATGTTG
	Dioon-cox2i2-R1848	CAACATAGTACAGTGCCTCACC
	Fouquieria-cox2i2-F890	CTGACCCTCTTCTTGTTCC
	Fouquieria-cox2i2-R908	GGAACAAGAAGAGGGTCAG
	Ginkgo-cox2i2-F471	CAATGAGAACACCTGACCCAACAG
	Ginkgo-cox2i2-F494	CTGTTGGGTCAGGTGTTCTCATTG

Magnoliako-cox2i2-R1858	CGTCAGAGTCCGTCAGTTCG
Neoregelia-cox2i2-F1202	TTAGTGTGAAAGGCTCAGGAA
Neoregelia-cox2i2-R1222	TTCCTGAGCCTTTCACACTAA
Oxalisreg-cox2i2-F1381	CGTGACTTCAGGAGAAGAGTG
Oxalisreg-cox2i2-F1476	GGCAGCATTGATAGACCGTTGAA
Oxalisreg-cox2i2-F570	CCAAGCAGAGCAACTAGAAGCC
Oxalisreg-cox2i2-R1401	CACTCTTCTCCTGAAGTCACG
Oxalisreg-cox2i2-R1498	TTCAACGGTCTATCAATGCTGCC
Oxalisreg-cox2i2-R572	GGCTTCTAGTTGCTCTGCTTGG
Persea-cox2i2-F1417	GCAAGGAAGGAGGCATTG
Persea-cox2i2-R1434	CAATGCCTCCTTCCTTGC
Pouteria-cox2i2-F916	GTCAAGATAGATGGGAAGGAG
Pouteria-cox2i2-R936	CTCCTTCCCATCTATCTTGAC
Ruscus-cox2i2-F1091	GTGACTAGGTCCCTCTGTTG
Ruscus-cox2i2-R1110	CAACAGAGGGACCTAGTCAC
Saxifraga-cox2i2-F1630	CTGTTCGCTTTCTCGCTTC
Saxifraga-cox2i2-R1500	CAACAACTACAAGGGTGCT
Zamia-cox2i2-F1471	GCAGTGGAGCTAACCTTGAATC
Zamia-cox2i2-F2168	GTATGGAACGGGTGCCTGA
Zamia-cox2i2-R1492	GATTCAAGGTTAGCTCCACTGC
Zamia-cox2i2-R2186	TCAGGCACCCGTTCCATAC
Zamiaflo-cox2i2-F1787	CTCTCACTCAGGACATAGC

Table S4. Primers used in this study, anchoring in the *cox2* intron2 (i691) of plants.

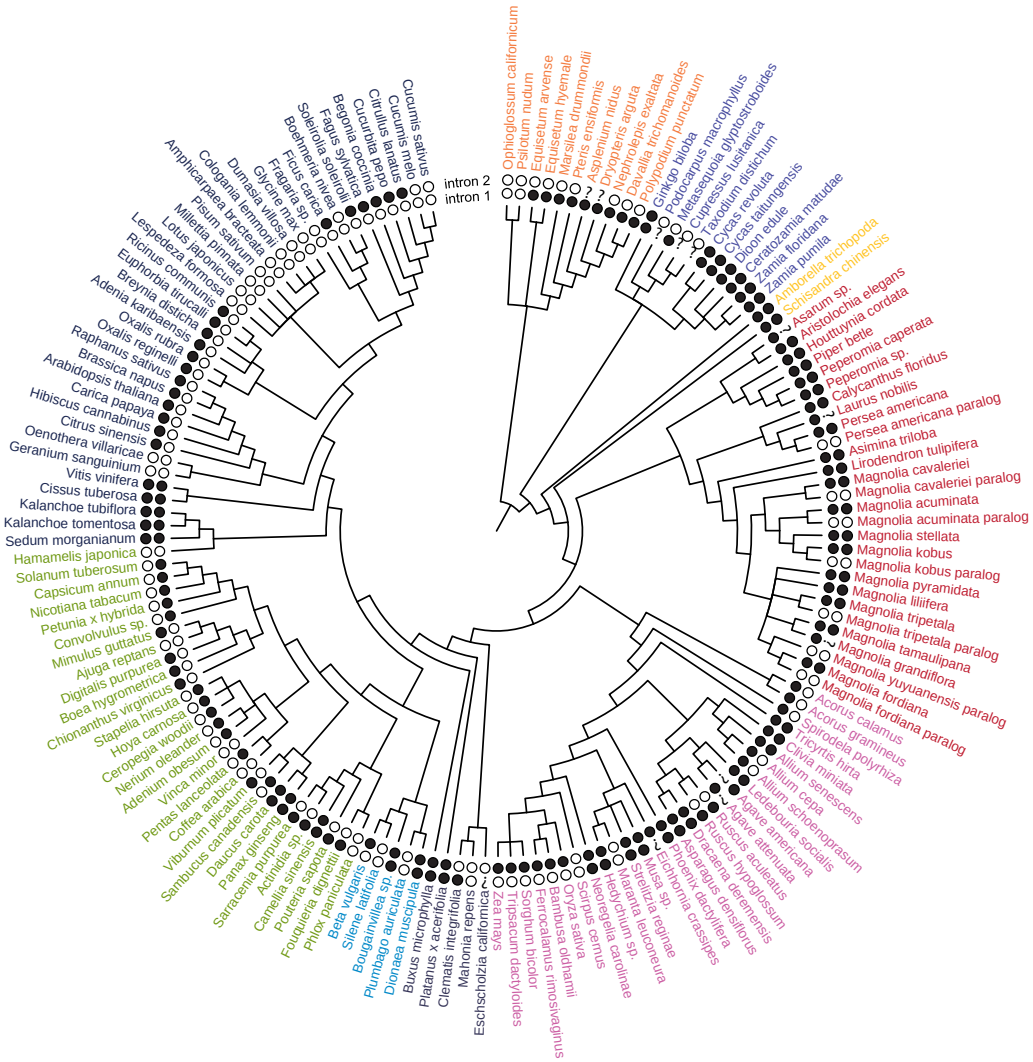
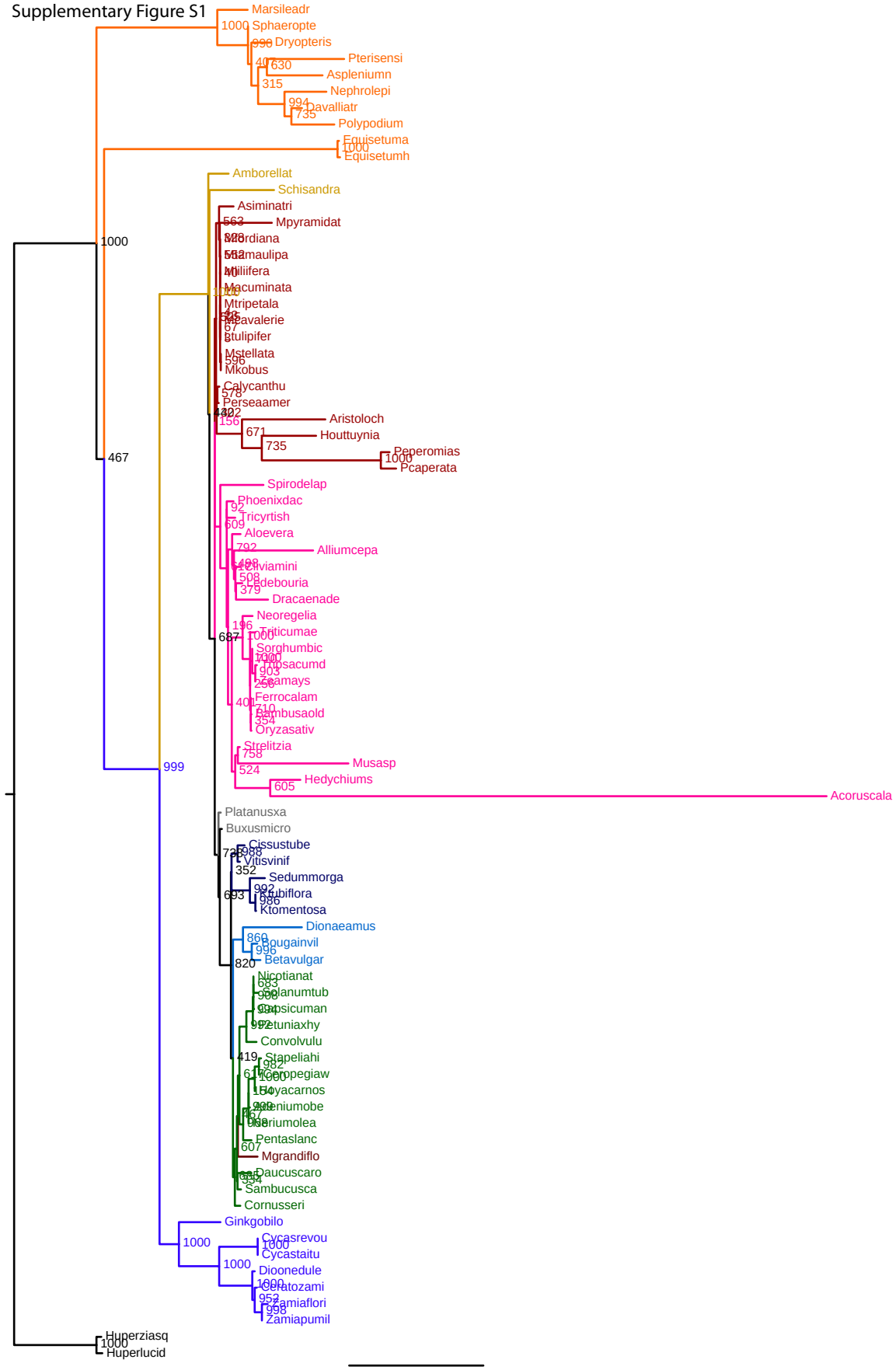
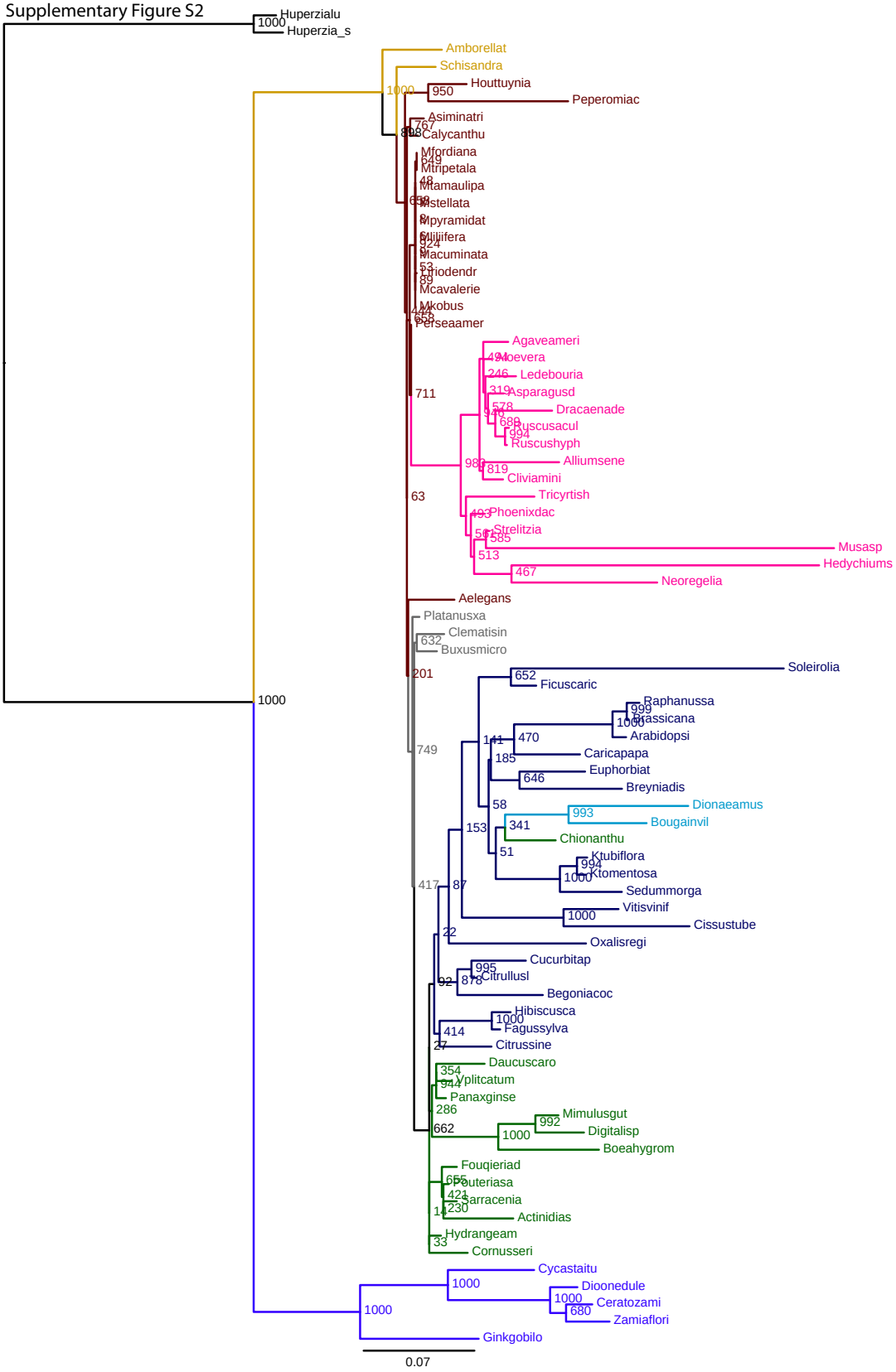


Figure S0. Phylogenetic distribution of the cox2 intron 1 (i373) and 2 (i691) across angiosperms, gymnosperms and ferns. Filled circles indicate intron presence while open circles indicate intron absence. Results not corroborated by PCR or sequencing are indicated as question marks. Species are color coded as following: monilophytes (orange), gymnosperms (blue), magnoliids (dark red), basal angiosperms (gold), monocots (magenta), basal eudicots (dim gray), rosids (navy), caryophyllids (turquoise blue), and asterids (forest green).

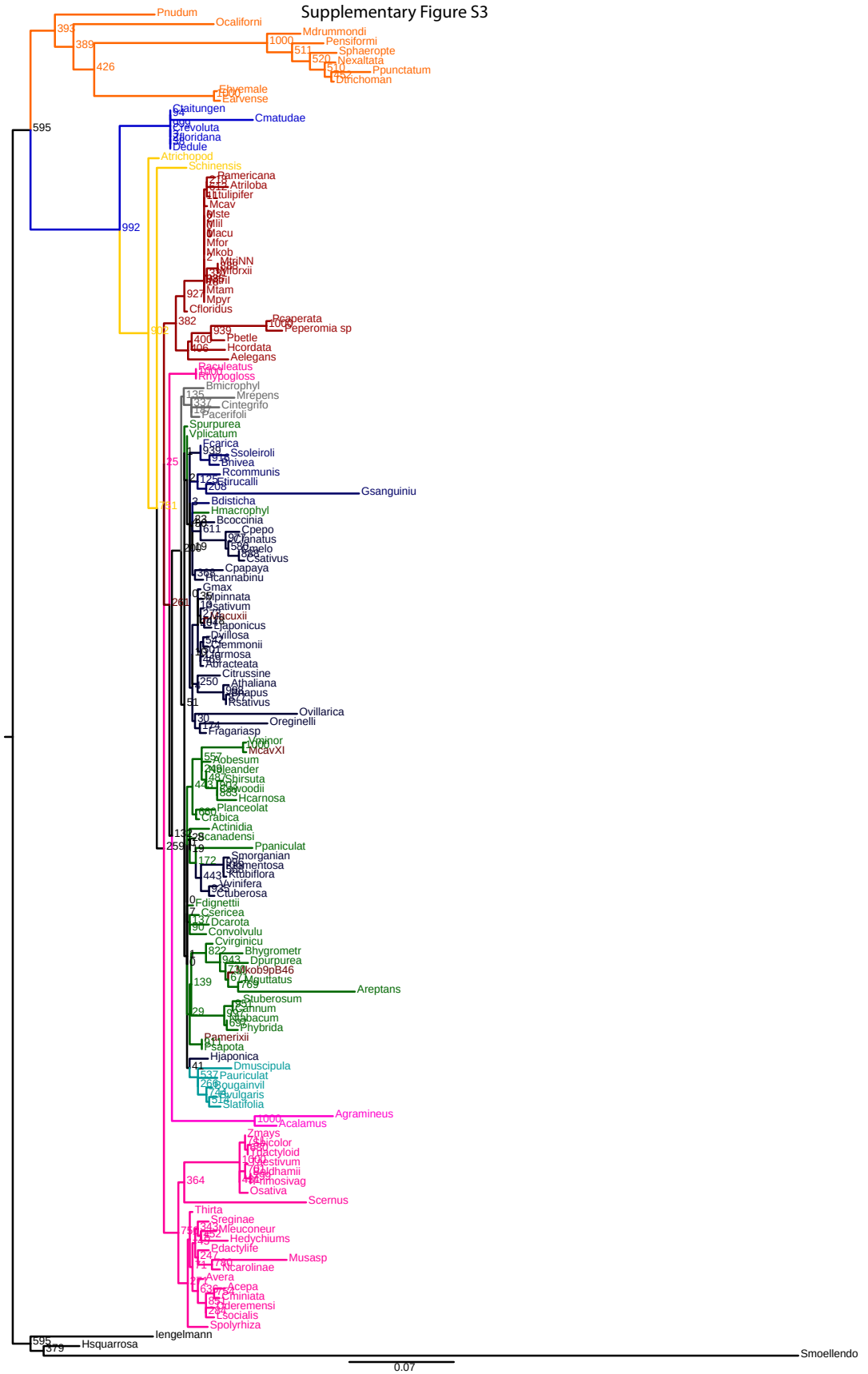
Supplementary Figure S1



Supplementary Figure S2



Supplementary Figure S3



Supplementary Figure S4